

Qy 2020 TCATGTTTAAAAA 2079
 Db 96 GAACCTTTAAAAA 37
 Qy 2080 AAAAAA 2106
 Db 36 AAAAAA 10

RESULT 4
 LOCUS BM884782
 DEFINITION rcl5a03.y1 Meloidogyne hapla egg pAMP1 v1 Meloidogyne hapla cDNA
 5', mRNA sequence.
 ACCESSION BM884782
 VERSION BM884782.1 GI:19268526
 KEYWORDS EST.
 SOURCE Meloidogyne hapla
 ORGANISM Meloidogyne hapla

REFERENCE 1 (bases 1 to 336)
 AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, F., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelis, V., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilton, R.

TITLE The Washington Univ. Nematode EST Project, 1999
 JOURNAL Unpublished
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 The library was constructed by Claire Murphy and Dr. James McCarter at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center
 Seq primer: -40RP from Gibco
 High quality sequence stop: 302.

FEATURES
 source
 1. .336
 /organism="Meloidogyne hapla"
 /mol_type="mRNA"
 /db_xref="taxon:6305"
 /dev_stage="parasitic adult females"
 /lab_host="DH10B"
 /clone_lib="Meloidogyne hapla egg pAMP1 v1"
 /note="Vector: pAMP1 (Gibco); Site.1: NotI; Site.2: SalI;
 The library was constructed by Claire Murphy and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dynal). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis kit from Clontech. Directionally cloned into the UDG sites of pAMP1."

BASE COUNT 207 a 21 c 31 g 77 t
 ORIGIN

Query Match 4.5%; Score 95.4; DB 12; Length 336;
 Best Local Similarity 72.8%; Pred. No. 9.8;
 Matches 123; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 1938 ATGTGTGATGTACGATAGCGGTGAAGAATTTTGTCTATCTACTGAACCTGCTATTGTC 1997
 Db 154 ATGTTCAATTTTATATATTTTAAAAAATTCATGTTTCACAAATTTCTGTTGTA 213

Qy 1998 TGAATCTTATGGCAGCTATGATCTATTTAAAAA 2057
 Db 214 TAAAAATTTTGGCTTTATTTGAAAAA 273

Qy 2058 AAAAAA 2106
 Db 274 AAAAAA 322

RESULT 5
 LOCUS AI392214/c
 DEFINITION NCSP19F7 Subtracted Perithecial Neurospora crassa cDNA clone SP19
 3' similar to mnt1 protein homolog - thiamine biosynthesis enzyme,
 mRNA sequence.
 ACCESSION AI392214
 VERSION AI392214.1 GI:4220021
 KEYWORDS EST.
 SOURCE Neurospora crassa
 ORGANISM Neurospora crassa

REFERENCE 1 (bases 1 to 480)
 AUTHORS Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L., Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, E., Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K., Miller, R., Ortega, J., Pavlova, I., Perea, J., Todisco, S., Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S. and Natvig, D.O.

TITLE Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa
 JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)
 MEDLINE 97435549
 PUBMED 9290248

COMMENT Contact: Natvig, D.O./Nelson, M.A.
 Department of Biology
 University of New Mexico
 Castetter Hall, Albuquerque, NM 87131, USA
 Tel: 505 277 3411
 Fax: 505 277 0304
 Email: ngp@biology.unm.edu

FEATURES
 source
 1. .480
 /organism="Neurospora crassa"
 /mol_type="mRNA"
 /strain="Fl a (FGSC 4347)"
 /db_xref="taxon:5141"
 /clone="SP19"
 /sex="Mating type a (fluffy), fertilized"
 /tissue_type="Perithecia"
 /dev_stage="Fruiting Body"
 /lab_host="E. coli"
 /clone_lib="Subtracted Perithecial"
 /note="Vector: pBlueScript SK (-); Site.1: EcoRI; Site.2: XhoI; mRNA isolated from 5 day old perithecia (fruiting bodies) of the fluffy strain fl a (Mating type a), type A). cDNA directionally cloned into pBlueScript SK(-) using the Uni-ZAP XR vector system (Stratagene, La Jolla, CA). Previously identified highly expressed clones were subtracted from this library."

BASE COUNT 128 a 137 c 48 g 166 t
 ORIGIN

Query Match 4.5%; Score 95.4; DB 9; Length 480;
 Best Local Similarity 67.2%; Pred. No. 7.9;
 Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 1906 ACTTCAAGTTGAATGATCTTCTGGTCCCTACCATGTTGATGTACCATAGCGGCTTGA 1965
 Db 209 AGTTCAGATGAGAGTTTGTGTGGATGATGGGCGCTCAGGGGTAGTAGGGAAG 150

Qy 1966 GATTTGTGTATTACTGAACCTGTACTTCTGCTGAATAGTTATGGCAGCTATGATTCATGT 2025
 Db 149 GGTTGGGTGAGTAAAGTTGTTTCCAGTTTAAATTTAGGTGAGTGTGTTTAAAAA 90

Qy 2026 TTAATAA 2085

[illegible]

RESULT 10	128	AAAAAAAAAAAAAAAAAAAAAAAAAAAA	154
BI746603			
LOCUS			
DEFINITION	rm29d01.y1	Meloidogyne arenaria egg pAMP1 v1 Chiapelli	McCart
ACCESSION	BI746603		
VERSION	BI746603.1	GI:15768405	
KEYWORDS	EST.		
SOURCE	Meloidogyne arenaria		
ORGANISM	Meloidogyne arenaria		
REFERENCE	1	(bases 1 to 165)	
AUTHORS	McCart, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.		
TITLE	The Washington Univ. Nematode EST Project, 1999		
JOURNAL	Unpublished		
COMMENT	Contact: McCart JP The Washington Univ. Nematode EST Project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu The library was constructed by Brandi Chiapelli and Dr. James McCart (bchiapelle@watson.wustl.edu & jmcarte@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. Nematodes were provided by Dr. David Bird and Daniel Snyder of North Carolina State University. Seq primer: -40RP from Gibco High quality sequence stop: 164.		
FEATURES	source		
	1. .165	Location/Qualifiers	
	/organism="Meloidogyne arenaria"		
	/mol_type="mRNA"		
	/db_xref="taxon:6304"		
	/dev_stage="egg"		
	/lab_host="DH10B"		
	/clone_lib="Meloidogyne arenaria egg pAMP1 v1 Chiapelli McCart"		
	/note="Vector: pAMP1 (Gibco); Site 1: Noti; Site 2: SalI; The library was constructed by Brandi Chiapelli and Dr. James McCart at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dt priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA synthesis kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Nematodes were provided by Dr. David Bird and Daniel Snyder of North Carolina State University."		
BASE COUNT	117 a	8 c 6 g	34 t
ORIGIN			
Query Match	4.5%;	Score 94.2;	DB 12; Length 165;
Best Local Similarity	77.6%;	Pred. No. 19;	
Matches 114;	Conservative	0; Mismatches	33; Indels 0; Gaps 0;
QY	1960	TTGAAGATTTTGTCTATTACGACCTGTACTTTGCTGTAATAGTTATGGCACTATGAT	2019
Db	8	TTTAAGATTTTGTCTATTACCTTTCGCTTAATTTTGAATAAACTGCTTTGAT	67
QY	2020	TCATCTGTTTAAAAA	2079
Db	68	AAAAA	127
QY	2080	AAAAA	2106

Db	128	AAAAAAAAAAAAAAAAAAAAAAAAAAAA	154
RESULT 11			
BI736597			
LOCUS			
DEFINITION	rk59f08.y1	Meloidogyne javanica egg pAMP1 v6 Chiapelli	McCart
ACCESSION	BI736597		
VERSION	BI736597.1	GI:14086286	
KEYWORDS	EST.		
SOURCE	Meloidogyne javanica	(root-knot nematode)	
ORGANISM	Meloidogyne javanica		
REFERENCE	1	(bases 1 to 286)	
AUTHORS	McCart, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.		
TITLE	The Washington Univ. Nematode EST Project, 1999		
JOURNAL	Unpublished		
COMMENT	Contact: McCart JP The Washington Univ. Nematode EST Project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu The library was constructed by Brandi Chiapelli and Dr. James McCart (bchiapelle@watson.wustl.edu & jmcarte@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. High quality sequence stop: 282.		
FEATURES	source		
	1. .286	Location/Qualifiers	
	/organism="Meloidogyne javanica"		
	/mol_type="mRNA"		
	/db_xref="taxon:6303"		
	/dev_stage="enriched for eggs"		
	/lab_host="DH10B"		
	/clone_lib="Meloidogyne javanica egg pAMP1 v6 Chiapelli McCart"		
	/note="Vector: pAMP1 (Gibco); The library was constructed by Brandi Chiapelli and Dr. James McCart at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dt priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA synthesis kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Nematodes were provided by Dr. David Bird of North Carolina State University."		
BASE COUNT	174 a	9 c 34 g	69 t
ORIGIN			
Query Match	4.5%;	Score 94.2;	DB 10; Length 286;
Best Local Similarity	77.6%;	Pred. No. 14;	
Matches 114;	Conservative	0; Mismatches	33; Indels 0; Gaps 0;
QY	1960	TTGAAGATTTTGTCTATTACGACCTGTACTTTGCTGTAATAGTTATGGCACTATGAT	2019
Db	116	TTGATATATTTTGTCTGTTGATTTAAAAATTTATTTATATAATTTTATATAATG	175
QY	2020	TCATCTTAAAAA	2079
Db	176	CTTTCTTAAAAA	235
QY	2080	AAAAA	2106
Db	236	AAAAA	262

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 01:59:13 ; Search time 133 Seconds

(without alignments)

6989.119 Million cell updates/sec

Title: US-09-856-327-1

Perfect score: 2106

Sequence: 1 atcagccatgtctctctca.....aaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.4	10.1	1869	3	US-09-305-381-1
2	210.2	10.0	1869	1	US-08-734-925-1
3	152.6	7.2	1701	3	US-09-023-731-2
4	91.6	4.3	2323	4	US-09-149-476-24
5	88.8	4.2	1733	3	US-09-073-569-1
6	88.4	4.2	1091	4	US-09-328-965-1
7	88	4.2	1248	4	US-09-489-847-101
8	88	4.2	2246	4	US-09-363-708-3
9	88	4.2	2246	4	US-09-083-587-3
10	87.2	4.1	2434	4	US-09-489-847-67
11	87	4.1	1114	4	US-09-152-060-41
12	87	4.1	2239	4	US-09-196-390-1
13	86.6	4.1	2671	6	5168051-9
14	85.6	4.1	1052	4	US-09-489-847-23
15	85	4.0	1582	3	US-08-545-196B-10
16	85	4.0	1582	3	US-08-545-196B-12
17	84.8	4.0	578	4	US-09-602-877A-95
18	84.8	4.0	3060	4	US-09-996-243-504
19	84.8	4.0	3715	4	US-09-234-245-1
20	84.8	4.0	7724	4	US-08-486-049-1
21	84.6	4.0	1037	4	US-09-489-847-112
22	84.6	4.0	1696	4	US-09-835-811-1
23	84.4	4.0	1249	4	US-09-461-325-128
24	84.4	4.0	1641	1	US-08-300-903A-8
25	84.4	4.0	1641	4	US-08-988-197-8
26	84.4	4.0	3410	3	US-09-020-956-110
27	84.4	4.0	3410	3	US-09-030-607-110

28	84.4	4.0	3410	4	US-09-439-313-110	Sequence 110, App
29	84.4	4.0	3410	4	US-09-352-616A-110	Sequence 110, App
30	84.4	4.0	3410	4	US-09-602-877A-100	Sequence 100, App
31	84.4	4.0	3410	4	US-09-232-149A-110	Sequence 110, App
32	84	4.0	441	4	US-09-601-537-10	Sequence 10, Appl
33	84	4.0	4121	4	US-09-601-537-9	Sequence 9, Appl
34	83.6	4.0	1474	3	US-08-821-994-64	Sequence 64, Appl
35	83.6	4.0	6409	4	US-09-967-908A-1	Sequence 1, Appl
36	83.4	4.0	1260	4	US-09-461-325-93	Sequence 93, Appl
37	83.4	4.0	1602	1	US-08-530-950-3	Sequence 3, Appl
38	83.4	4.0	1602	3	US-08-888-429A-3	Sequence 3, Appl
39	83.4	4.0	1602	3	US-09-149-879-3	Sequence 3, Appl
40	83.4	4.0	1602	4	US-09-057-009-3	Sequence 3, Appl
41	83.4	4.0	1895	4	US-09-444-336-7	Sequence 7, Appl
42	83.2	4.0	674	4	US-09-620-405B-465	Sequence 465, App
43	83.2	4.0	674	4	US-09-433-826B-465	Sequence 465, App
44	83.2	4.0	674	4	US-09-604-287A-465	Sequence 465, App
45	83.2	4.0	1361	4	US-09-489-847-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1

US-09-305-381-1

; Sequence 1, Application US/09305381

; Patent No. 6146865

; GENERAL INFORMATION:

; APPLICANT: Christensen, Soren

; APPLICANT: Lassen, Soren Flensted

; APPLICANT: Schneider, Palle

; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides

; TITLE OF INVENTION: Having Pyranose Oxidase Activity

; FILE REFERENCE: 5571.200-US

; CURRENT APPLICATION NUMBER: US/09/305.381

; CURRENT FILING DATE: 1999-05-05

; EARLIER APPLICATION NUMBER: 60/088,724

; EARLIER FILING DATE: 1998-06-10

; EARLIER APPLICATION NUMBER: PA 1998 00774

; EARLIER FILING DATE: 1998-06-08

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 1869

; TYPE: DNA

; ORGANISM: Trameetes hirsuta

US-09-305-381-1

Query Match	10.1%;	Score 212.4;	DB 3;	Length 1869;
Best Local Similarity	51.5%;	Pred. No. 6.7e-39;		
Matches	841;	Conservative	0;	Mismatches 656;
				Indels 135;
				Gaps 10;
Qy	269	GGCTACACACAGAAGATGAATCGAGTTCGAGAAAGATATTGACCGCTTCGTCATGTGA 328		
Db	259	GGCTACACACAGAAGAACACCGTCGAGTACCAAGAAACATCGACAAATTCGTAATGTT 318		
Qy	329	ATCAAGGCGCTTACACAAAGTCTCTGTTCTCTGAGAACACAGAGCTGCTTACACTT 388		
Db	319	ATCAAGGCGCAACTTATGCCGCTCGTCCGCTCAACACGATGGTGTGACACGCTA 378		
Qy	389	GATCCCGGAGCTGGAGCGCGCCCTTGAAGTTTACGCCATPATCAACGGTFAAAATTCCT 448		
Db	379	AGCCCGCGCTATGGCAAGCTTTCGACG-----TTCCTCGTCCGCAACGGGGCGAATCCA 432		
Qy	449	CACCAGCGGAATTCGAGAACTTCTCTCGGAGCGCGTCAAGCGTGGAGTCGGCGCATG 508		
Db	433	GAGCAAGACCGCTGCGCAACTTATGGCGAGCGGTACCCCGCTCGTGGCGCATG 492		
Qy	509	AGTACCACCTGGAGCTGCTCCACGCGACGGATTTCATCCACCCATGCGAAAGTCTCCCGGGC 568		
Db	493	TCTACGACCTGGAGCTGCGGACCGCGCTTCGAGAAAGTCTCCCGGGC----- 537		
Qy	569	ATCGGCGCTCCGAAGCTCAGTAACGACCGCGCAGAGGACGACAAAGAGTGAACGAGCTT 628		

Best Local Similarity 50.4%; Pred. No. 2.le-38;
Matches 857; Conservative 0; Mismatches 713; Indels 129; Gaps 9;

```
QY 200 GAGATCGGAGTGTGATAGTCTTACGCTGTTAAATGCCGAAGAAGAACTGCAGTTCCC 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 GAGCTCTGCAAGCCGGTTACAGGTCCCATGTTTCACATCGGGAAATTTGACTTGGC 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 260 TAGCTTCTCGTACCAACAAGAAGAAATGAATCGAGTTCCAGAAAGATATTGACCGCTTC 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 CTGAAGATCGTGGCCCAACAAGAACAACACCGCTCGAATACCAAGAACAATGACAAGTTT 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 320 GTCAATGTAATCAAGGAGCGCTTACACAAGTCTCTGTTCTGTCAGAAACCAAGAGTG 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 GTGAACGTCATTCAGGCGCAATATGATGTCGTCTGTTTCCGTTCCCGCTCAATACCCCTGTC 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 380 CTTACACTTGTATCCCGGAGCGCTGGAGCGCCCTCGAAGTTCAGCCATATCGAACGGT 439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GACACGCTCAGCCGAGCGTCTTGGCAAGCTTCATCG-----TTCCTCGTCCGCAATGGC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 440 AAAAATCCTCACCAGCGGGAATTGAGAACTTGTCTCGGAGGCCGTAACGCGTGGAGTC 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 TCGAAACCCAGAGGAGGACCCGCTTCGTAACCTCAGTGTCTCAGCGGTCACCGCTGTCGTC 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 500 GCGGGATGATACCCACTGAGCTGCTCCACGCCACGGATTATCCACCCATGGAAAGT 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 GGAGGCATGTCCACGCACTGGACATGCGGACACCGCGCTTGA----- 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 560 CTCGCGGCACTCGCCGCTCGAAGCTCAGTAACGCCCGCAGAGGACGACAAAGATGG 619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 --CGCGAGGAGCGCCGCTGCTGCTGAAGACACCCAGACGCTGACGAGCGGAGTGG 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 620 AAGAGCTTTATTCGAGGCGGAGCGTCTCATCGGACTTCCACCAAGGAATTCGACGAG 679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 GACCGGCTGTACACCAAGGCGGAGTCACTTCAAGACCAGGACCGGACGATTCAGGAG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 680 TCAATTCGGACACACCTTGTCTGCGCTCTTTCGAACGCGGTACAGGATTCGTCACGT 739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 TCGATCGGCCACACCTCGTGTCTCAACAACTCGCGGAGGAATACAAAGTTCAGCGGCAC 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 740 ATCTTTTCGCCCTCTCCCGTTGGCATGCGACCGTGAAGAACGCGGGAATACGTCGAA 799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 TTCAGAGATCCCGCTCGCGGCAACGCGTCG-----CAGTCGACCTTCGTCGAG 756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 800 TGGCACTCAGAGAAATCTTTTCCACTCTATCTACACGATGACAGAGAGAGCTC 859
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 TGGAGCTCGGCGAACAACCGTGT-----CGACCTCCAGACAGCGCCGAAC 801
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 860 TTTACCCTGTGACGAACCATCGCTGCACAGACTGGCGCTTACGGCGGGTATGAGAAG 919
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 ACGAGCGCGGGAATGAGCGCTTCAACCTCTTCCCGCGGTTGCAATGTGAGCGCGTCGTG 861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 920 AAGATTGGCGTGGCGAGGTGAGGAATCTACTGGCCACAGGAATCTAGTTTCGACGCTG 979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 862 CGCAACAGCTCGAATCCGAGATCGAGAGTCTGCACATCACGACCTCATCTCGGGCGAC 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 980 GACAGCTATATCATGCGAAGGTATATGTAAGTGGCGTGGGAGCGATCGGCAACCCACAG 1039
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 922 CGCTTCAAAATCAAGACAGAGCTGTTCTGTTTACAGCCGGGCGGTCCCAACAGCGCAG 981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1040 ATTCTCTATACCTGGCGCTT-----CTCTGGGTACAGTCAAGCCACCGCAATGACTCG 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 982 CTTCTGCTGAACTCTGGCTTTGGACAGCTGGCGCGCGGACCCCGGAAACCCGCGCAG 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1094 TTGATCCCAACCTGGGAGGTATACATCAGGAGCGCGATGGCATTTTGCAGATAGTC 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1042 TTGCTGCGCTCCCTGGGAAGCTATATCAGGACGATCGCTGCTCTCTCCGACCGCTG 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1154 TTGAGGAGGAATTCGTACAGCGTG-----CGCGAGATCTCTTATGAGCTGCC----- 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1102 ATGAGCCGAGCTCATCGACAGCGTCAAGTCCGACATGATCATCAGGGGCAACCCCTGGC 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1204 -----ATGC 1207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 3

US-09-023-731-2

; Sequence 2, Application US/09023731

; Patent No. 6291648

; GENERAL INFORMATION:

; APPLICANT: Kawamura, Yukio; Morita,

; APPLICANT: Akhiro; Izumo, Koji.; Saka, Tomohide.

; TITLE OF INVENTION: ANTITUMOR PROTEIN AND

; TITLE OF INVENTION: GENE ENCODING SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD 97

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,731

; FILING DATE: 13-FEB-1998

; PRIOR APPLICATION DATA:

RESULT 9
US-09-083-587-3
: Sequence 3, Application US/09083587
: Patent No. 6492138
: GENERAL INFORMATION:
: APPLICANT: Schmandt, et al.
: TITLE OF INVENTION: NOVEL SHC BINDING PROTEIN
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 233 South Wacker Drive/6300 Sears Tower
: CITY: Chicago
: STATE: Illinois

[illegible]

```

DB          2228 AAAAAAAAAAAAAAAAAA 2243

RESULT 10
US-09-489-847-67
; Sequence 67, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657

```

; EARLIER FILING DATE: 1998-07-30
 ;
 ; EARLIER APPLICATION NUMBER: 60/095,486
 ;

; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 67
; LENGTH: 2434
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (75)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (103)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (130)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-489-847-67

Query Match 4.1%; Score 87.2; DB 4; Length 2434;
Best Local Similarity 76.4%; Pred. No. 1.4e-10;
Matches 107; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
Qy 1967 ATTTTGTCTATTACTGAACCTGTACTTGTCTGAATAGTATGTCACCTATGATTCATGTT 2026
Db 2292 ATATTTTTTTTAAAGTAAATGGACCCAGTAAGAAATTAATAATACCAACATATAAAA 2351
Qy 2027 TAAAAAATAAA 2086
Db 2352 AAA 2411
Qy 2087 AAAAAAAAAAAAAAAAAAAAAA 2106
Db 2412 AAAAAAAAAAAAAAAAAAAAAA 2431

RESULT 11
US-09-152-060-41
; Sequence 41, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762

; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-152-060-41

Query Match 4.1%; Score 87; DB 4; Length 1114;
Best Local Similarity 77.9%; Pred. No. 1.2e-10;
Matches 102; Conservative 2; Mismatches 27; Indels 0; Gaps 0;
Qy 1976 ATTACTGAACCTGTACTTGTCTGAATAGTATGTCACCTATGATTCATGTTTAAAAA 2035
Db 981 ATTATGGCACCCTGAATTAGGACAGTGCATTAATAAGTTGGCTKTTTAAWATTTTAAAA 1040
Qy 2036 AAA 2095
Db 1041 AAA 1100
Qy 2096 AAAAAAAAAAAAAA 2106
Db 1101 AAAAAAAAAAAAAA 1111

RESULT 12
US-09-196-390-1
; Sequence 1, Application US/09196390
; Patent No. 6307125
; GENERAL INFORMATION:
; APPLICANT: Block, Martina
; APPLICANT: Lorz, Horst
; APPLICANT: Luticke, Stephanie
; APPLICANT: Walter, Lennart
; APPLICANT: Froberg, Claus
; APPLICANT: Kossmann, Jens
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ENZYMES
; TITLE OF INVENTION: FROM WHEAT WHICH ARE INVOLVED IN STARCH
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,390
; FILING DATE:
; CLASSIFICATION:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: DE 196 21 588.9
;; FILING DATE: 29-MAY-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: DE 196 36 917.7
;; FILING DATE: 11-SEP-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/EP97/02793
;; FILING DATE: 28-MAY-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Haley, Jr., James F.
;; REGISTRATION NUMBER: 27,794
;; REFERENCE/DOCKET NUMBER: AGREVO-9
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 596-9000
;; TELEFAX: (212) 596-9090
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2239 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to mRNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Triticum aestivum L.
;; STRAIN: cv. Florida
;; HAPLOTYPE: ca. 21 d Caryopses
;; IMMEDIATE SOURCE:
;; LIBRARY: cDNA library in pBluescript sk (-)
;; CLONE: TaSSS
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..2017
;;
US-09-196-390-1

Query Match 4.1%; Score 87; DB 4; Length 2239;
Best Local Similarity 67.2%; Pred. No. 1.5e-10;

Matches 123; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1924 TCTGGTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
DB 2054 TGTGTTGTTGCTTGTAGCTGACAAATATTGACCTGTTGGAGAAATTTTATCTTGTGCT 2113
QY 1984 ACCTGTACTTTGCTGAATAGTTATGGCACTATGATTCATGTTTAAAAAATAAAAAA 2043
DB 2114 GTTTTAAATCAAGAGGGGTTTCTCCGATTTTCATTAAAAAATAAAAAA 2173
QY 2044 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2103
DB 2174 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2233
QY 2104 AAA 2106
DB 2234 AAA 2236

RESULT 13

5168051-9
; Patent No. 5168051
; APPLICANT: DERYNCK, RIK M.A.; GOEDDEL, DAVID V.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/389,929
; FILING DATE: 04-AUG-1989
; SEQ ID NO: 9:
; LENGTH: 2671
5168051-9

Query Match 4.1%; Score 86.6; DB 6; Length 2671;
Best Local Similarity 83.8%; Pred. No. 2e-10;

Matches 98; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 1990 ACTTGTCTGTAAGTATGGCACTATGATTCATGTTTAAAAAATAAAAAAATAAAAAA 2049
DB 2544 ATTGTTCTAAACAATAAAGCGCTTATTCTAGGTGTAATAAAAAAATAAAAAAATAAAAAA 2603
QY 2050 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2106
DB 2604 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2660

RESULT 14

US-09-489-847-23
; Sequence 23, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031PI
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1052
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-23

Query Match 4.1%; Score 85.6; DB 4; Length 1052;
Best Local Similarity 95.7%; Pred. No. 2.5e-10;

Matches 88; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2015 ATGATTTCATGTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2074
DB 954 ATTTTAAATTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1013
QY 2075 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2106
DB 1014 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1045

RESULT 15

US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-545-196B-10

Query Match 4.0%; Score 85; DB 3; Length 1582;
Best Local Similarity 94.6%; Pred. No. 3.8e-10;
Matches 88; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 2014 TATGATTCATGTTTAAA 2073
Db 1477 TTTAATTTT TTTTAAA 1536
Qy 2074 AA 2106
Db 1537 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1569

Search completed: September 24, 2003, 05:33:21
Job time : 143 secs


```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701
; TYPE: nucleic acid
; TOPOLOGY: linear
; (D) STRANDNESS: Single
; MOLECULE TYPE: cDNA to RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-911-969-2

Query Match
Best Local Similarity 7.2%; Score 152.6; DB 10; Length 1701;
Matches 394; Conservative 0; Mismatches 304; Indels 39; Gaps 2;

QY 1106 CTGGGAGGTACATCAGGAGCAGCCGATGTCATTTCCAGATAGTCTTGAGGAGGAA 1165
Db 1106 CTGGGAGGTACATCAGGAGCAGCCGATGTCATTTCCAGATAGTCTTGAGGAGGAA 1165
QY 964 CTGGTGTCTACCTCAGCGAACAGTCCATGACTTTTTTGTGATGCTTCAAGAGGGGC 1023
Db 964 CTGGTGTCTACCTCAGCGAACAGTCCATGACTTTTTTGTGATGCTTCAAGAGGGGC 1023
QY 1166 TTCGTGACAGGTCGGCGAGATCCTTATGAGTGCATGTTGGAAAGAGCCGTTGCT 1225
Db 1166 TTCGTGACAGGTCGGCGAGATCCTTATGAGTGCATGTTGGAAAGAGCCGTTGCT 1225
QY 1024 ATAGTCGATGCCATCGCTACTGACCT-----CGCTTCGCTCGGAAGTTGAG 1071
Db 1024 ATAGTCGATGCCATCGCTACTGACCT-----CGCTTCGCTCGGAAGTTGAG 1071
QY 1226 CAACATATTGCAAGAACCCGACAGATGCATGCGCATTCGTTCCGCGATCCGGAACCC 1285
Db 1226 CAACATATTGCAAGAACCCGACAGATGCATGCGCATTCGTTCCGCGATCCGGAACCC 1285
QY 1072 GCGACACAGAGAACCCCGATGACGTGTGCGCCATTCATCCAGAGCTGAACCT 1131
Db 1072 GCGACACAGAGAACCCCGATGACGTGTGCGCCATTCATCCAGAGCTGAACCT 1131
QY 1286 CAGGTAAACCCCATTTACAGAGAACACCCCTGGCACAGCAGATTCACCGCATGCT 1345
Db 1286 CAGGTAAACCCCATTTACAGAGAACACCCCTGGCACAGCAGATTCACCGCATGCT 1345
QY 1132 CAAGTGATGATTCCTACACGTCGACCTTCCTTGGCATGTTTCAGGTGCATCGCGATGCA 1191
Db 1132 CAAGTGATGATTCCTACACGTCGACCTTCCTTGGCATGTTTCAGGTGCATCGCGATGCA 1191
QY 1346 TTTCTGACGTGCGCTCGGTCCTGAGTGACTCTCGTGCATGCTGACGTCGCGCTGG 1405
Db 1346 TTTCTGACGTGCGCTCGGTCCTGAGTGACTCTCGTGCATGCTGACGTCGCGCTGG 1405
QY 1192 TTCATATATGTGATGTTGGACCCAAAGCCGCGTGTGTGCTGATCTGAGGTTT 1251
Db 1192 TTCATATATGTGATGTTGGACCCAAAGCCGCGTGTGTGCTGATCTGAGGTTT 1251
QY 1406 TTTGGCGCAACCGCCCTGAAGCAACACCTTTTGGTTTC----- 1447
Db 1406 TTTGGCGCAACCGCCCTGAAGCAACACCTTTTGGTTTC----- 1447
QY 1252 TTCGCAATCAGATATTGTCGAGAAATCGAGTACTTTCGGTCCGAAACCTTAAGCTA 1311
Db 1252 TTCGCAATCAGATATTGTCGAGAAATCGAGTACTTTCGGTCCGAAACCTTAAGCTA 1311
QY 1448 -----CAGAACGATGTTCAAGACGGGTACAGTATCCGCGACCGACGTTTCAGATAT 1498
Db 1448 -----CAGAACGATGTTCAAGACGGGTACAGTATCCGCGACCGACGTTTCAGATAT 1498
QY 1312 CCGGAGTGGGAAGCGGGTGTACAGACACTTATGGAATGCCACAGCCGACATTCCTATC 1371
Db 1312 CCGGAGTGGGAAGCGGGTGTACAGACACTTATGGAATGCCACAGCCGACATTCCTATC 1371
QY 1499 CGACCCACACTGCTCAACAGTGAGACAGAGGAAATGATGGCGGATATGCGCAAGTG 1558
Db 1499 CGACCCACACTGCTCAACAGTGAGACAGAGGAAATGATGGCGGATATGCGCAAGTG 1558
QY 1372 AAGCGGACCAACCGCGATGAGACCGTGACACGAGGATGATGAATGATGACCAACGTC 1431
Db 1372 AAGCGGACCAACCGCGATGAGACCGTGACACGAGGATGATGAATGATGACCAACGTC 1431
QY 1559 GCGAGCAACTTGGGAGGTTATTGGCCACGCTCCCGCCGAGTTTATGGATCCAGGCGCT 1618
Db 1559 GCGAGCAACTTGGGAGGTTATTGGCCACGCTCCCGCCGAGTTTATGGATCCAGGCGCT 1618
QY 1432 CGAATATGCTGGTGGTACTCTTCTGCTCTACCTCAATTTATGGCACTGCTGCTC 1491
Db 1432 CGAATATGCTGGTGGTACTCTTCTGCTCTACCTCAATTTATGGCACTGCTGCTC 1491
QY 1619 GCATTCATCTTGGGGGACTACTCGCATTTGGCTTCGACAGGCAACTACAGTGGCTGAT 1678
Db 1619 GCATTCATCTTGGGGGACTACTCGCATTTGGCTTCGACAGGCAACTACAGTGGCTGAT 1678
QY 1492 GTACTGCACATCAGCGGAACCTACTCGGATCGGACAGATGATCAAACTTCTGTTGCTGAT 1551
Db 1492 GTACTGCACATCAGCGGAACCTACTCGGATCGGACAGATGATCAAACTTCTGTTGCTGAT 1551
QY 1679 AACAACTCGCTGGCTGGGACTTTGCCAATCTTTATGTTGAGGCAATGGCACCATCAGG 1738
Db 1679 AACAACTCGCTGGCTGGGACTTTGCCAATCTTTATGTTGAGGCAATGGCACCATCAGG 1738
QY 1552 CCGACATCAAGAGTTCATATCTCAACAACTGTGGGTCGCGGGAATGGGTGCAATCCA 1611
Db 1552 CCGACATCAAGAGTTCATATCTCAACAACTGTGGGTCGCGGGAATGGGTGCAATCCA 1611
QY 1739 ACGGCTTCGCGGAGAACCCGACACTTACGTCGATGTGCCAGCTTATCAAGAGCGCGAGG 1798
Db 1739 ACGGCTTCGCGGAGAACCCGACACTTACGTCGATGTGCCAGCTTATCAAGAGCGCGAGG 1798
QY 1612 GATGGACTGCTGCAACCCGACTGCTGACGAGCGTTCGCGTATGCGCTCAAGGGTCTGAG 1671
Db 1612 GATGGACTGCTGCAACCCGACTGCTGACGAGCGTTCGCGTATGCGCTCAAGGGTCTGAG 1671
QY 1799 AGCATCATCAATACACT 1815
Db 1799 AGCATCATCAATACACT 1815
QY 1672 GCTGTAGTCAATACCT 1688
Db 1672 GCTGTAGTCAATACCT 1688
```

RESULT 2
US-10-199-672-75
; Sequence 75, Application US/10199672
; Publication No. US2003014842A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian

```
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430RIC1  
; CURRENT APPLICATION NUMBER: US/10/199,672  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: US/10/052,586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 75  
; LENGTH: 4640  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-199-672-75

Query Match 4.5%; Score 94.4; DB 12; Length 4640;  
Best Local Similarity 72.9%; Pred. NO. 2e-09;  
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;  
QY 1924 TCTGTCCTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAGAGATTTTGTGTATTACTGA 1983  
Db 4423 TATGGCTCCATTTTATTTATAGTGTAAAGTTGTATTTCTTAAAGTTTGTGTTTGTGCGA 4482  
QY 1984 ACCTGCTACTTGTCTCGA-----ATAGTTATGGCAGCTATGATTCATGTTTAAAAA 2038  
Db 4483 CAGTATCTTTTAAATGAGCTCTTAAAAATAAAGGCATATGTTTCATGTTTAAAAA 4542  
QY 2039 AAAAAA 2098  
Db 4543 AAAAAA 4602  
QY 2099 AAAAAA 2106  
Db 4603 AAAAAA 4610
```

RESULT 3
US-10-187-749-75
; Sequence 75, Application US/10187749
; Publication No. US20030153036A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.


```
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,Victor
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC1
; CURRENT APPLICATION NUMBER: US/10/187,749
; CURRENT FILING DATE: 2002-07-01
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/052,63
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/052,66
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062,250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063,120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063,121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063,486
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063,540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063,541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063,544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-187-749-75

Query Match          4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGTCAGATAGCGGTTGAAAGATTTTGTGTTACTGCA 1983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4423 TATGCTCCATTTTATTTATAGTGTAAAGTTGTTTCTTAAAGTTTGTGTTGTCGA 4482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1984 ACCTGTACTTTGTCCTGA-----ATAGTTATGCGACTATGATTCATGTTTAAAAA 2038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4483 CAGTATCTTTTAAATGAGTCTTAAATAAAGGCATATTTGTTTCATGTTTAAAAA 4542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2099 AAAAAAAA 2106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4603 AAAAAAAA 4610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-10-194-457-75
; Sequence 75, Application US/10194457
; Publication No. US20030153037A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, Victor
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
```

```
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC296
; CURRENT APPLICATION NUMBER: US/10/194,457
; CURRENT FILING DATE: 2002-07-11
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/052,63
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/052,66
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062,250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063,120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063,121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063,486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063,540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063,541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063,544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-194-457-75

Query Match          4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGTCAGATAGCGGTTGAAAGATTTTGTGTTACTGCA 1983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4423 TATGCTCCATTTTATTTATAGTGTAAAGTTGTTTCTTAAAGTTTGTGTTGTCGA 4482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1984 ACCTGTACTTTGTCCTGA-----ATAGTTATGCGACTATGATTCATGTTTAAAAA 2038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4483 CAGTATCTTTTAAATGAGTCTTAAATAAAGGCATATTTGTTTCATGTTTAAAAA 4542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2099 AAAAAAAA 2106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4603 AAAAAAAA 4610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-10-184-642-75
; Sequence 75, Application US/10184642
; Publication No. US20030157635A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
```


Db	4423	TATGGCTCCATATTTATTTATATAGTGTAAAGTTGTATTTCTCTAAAGTTTGTGTTTGTGCGA	4482
Qy	1984	ACCTGTACTTTGTCTGTA-----ATAGTTATGGCACATATGATTCATGTTTAAAAA	2038
Db	4483	CAGTATCTTTTAAATGAGTCTTAAAAATAAAGCATATTCCTCATGTTTAAAAA	4542
Qy	2039	AAAAA	2098
Db	4543	AAAAA	4602
Qy	2099	AAAAAAA 2106	
Db	4603	AAAAAAA 4610	

RESULT 10
US-10-173-692-75
: Sequence 75, Application US/10173692
: Publication No. US20030166188A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Jian
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3430R1C20
: CURRENT APPLICATION NUMBER: US/10/173, 692
: CURRENT FILING DATE: 2002-06-17

```

; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-692-75

Query Match      4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGCTGTACGATAGCGGTGAAAGATTTTGTGTATTACTGA 1983
Db 4423 TATGGCTCCATTATTTTATAGTGTAAAGTTGTATTTCCTAAAGTTGTGTGTTTGTCTGA 4482

QY 1984 ACCTGTACTTTGTCCTGA-----ATAGTTATGGCACATATGATTCATGTTTAAAAAATAAAA 2038
Db 4483 CAGTATCTTTTAATGAGTCTTAAAAATAAAGGCATATGTTCTATGTTTAAAAAATAAAA 4542

QY 2039 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2098
Db 4543 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 4602

QY 2099 AAAAAAAA 2106
Db 4603 AAAAAAAA 4610

RESULT 11
US-10-173-694-75
; Sequence 75, Application US/10173694
; Publication No. US20030166107A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc

```

```
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C19
; CURRENT APPLICATION NUMBER: US/10/173,694
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-694-75

Query Match          4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGTCCTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4423 TATGGCTCCATTTTATTATTATAGTGTAAAGTTGTATTCCCTAAAGTTTGTGTTCGA 4482

QY 1984 ACCTGTACTTGTCTGA-----ATAGTTATGCGCACTATGATTCATGTTTAAAAA 2038
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4483 CAGTATCTTTTAAATGAGCTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602

QY 2099 AAAAAAAA 2106
      ||||| |||
Db 4603 AAAAAAAA 4610

RESULT 13
US-10-173-699-75
; Sequence 75, Application US/10173699
; Publication No. US20030166109A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C8
; CURRENT APPLICATION NUMBER: US/10/173,699
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-699-75

Query Match          4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGTCCTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4423 TATGGCTCCATTTTATTATTATAGTGTAAAGTTGTATTCCCTAAAGTTTGTGTTCGA 4482

QY 1984 ACCTGTACTTGTCTGA-----ATAGTTATGCGCACTATGATTCATGTTTAAAAA 2038
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4483 CAGTATCTTTTAAATGAGCTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602

QY 2099 AAAAAAAA 2106
      ||||| |||
Db 4603 AAAAAAAA 4610

RESULT 12
US-10-173-698-75
; Sequence 75, Application US/10173698
; Publication No. US20030166108A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C12
; CURRENT APPLICATION NUMBER: US/10/173,698
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-698-75

Query Match          4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
```

```
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGTCCTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4423 TATGGCTCCATTTTATTATTATAGTGTAAAGTTGTATTCCCTAAAGTTTGTGTTCGA 4482

QY 1984 ACCTGTACTTGTCTGA-----ATAGTTATGCGCACTATGATTCATGTTTAAAAA 2038
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4483 CAGTATCTTTTAAATGAGCTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602

QY 2099 AAAAAAAA 2106
      ||||| |||
Db 4603 AAAAAAAA 4610

RESULT 13
US-10-173-699-75
; Sequence 75, Application US/10173699
; Publication No. US20030166109A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C8
; CURRENT APPLICATION NUMBER: US/10/173,699
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-699-75

Query Match          4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGTCCTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4423 TATGGCTCCATTTTATTATTATAGTGTAAAGTTGTATTCCCTAAAGTTTGTGTTCGA 4482

QY 1984 ACCTGTACTTGTCTGA-----ATAGTTATGCGCACTATGATTCATGTTTAAAAA 2038
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4483 CAGTATCTTTTAAATGAGCTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602

QY 2099 AAAAAAAA 2106
      ||||| |||
Db 4603 AAAAAAAA 4610

RESULT 14
US-10-173-707-75
; Sequence 75, Application US/10173707
; Publication No. US20030166110A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C17
; CURRENT APPLICATION NUMBER: US/10/173,707
; PRIOR FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-707-75

Query Match      4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
Db 4423 TATGGCTCCATTTTATTTTATAGTGTAAAGTTGATTTCCTAAAGTTTGTGTTTGTGCGA 4482

QY 1984 ACCTGTACTTTGTCTGA-----ATAGTTTATGGCAGCTATGATTCATGTTTAAAAA 2038
Db 4483 CAGTATCTTTTAAATGAGTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602

QY 2099 AAAAAAAA 2106
Db 4603 AAAAAAAA 4610

Search completed: September 24, 2003, 05:56:49
Job time : 1401 secs

```

```

US-10-174-569-75

Query Match      4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
Db 4423 TATGGCTCCATTTTATTTTATAGTGTAAAGTTGATTTCCTAAAGTTTGTGTTTGTGCGA 4482

QY 1984 ACCTGTACTTTGTCTGA-----ATAGTTTATGGCAGCTATGATTCATGTTTAAAAA 2038
Db 4483 CAGTATCTTTTAAATGAGTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602

QY 2099 AAAAAAAA 2106
Db 4603 AAAAAAAA 4610

```

Search completed: September 24, 2003, 05:56:49
Job time : 1401 secs

```

RESULT 15
US-10-174-569-75
; Sequence 75, Application US/10174569
; Publication No. US2003016611A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C39
; CURRENT APPLICATION NUMBER: US/10/174,569
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien

```

```

US-10-174-569-75

Query Match      4.5%; Score 94.4; DB 12; Length 4640;
Best Local Similarity 72.9%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

QY 1924 TCTGGTCCCTACCATGTTGATGTGTACGATAGGCGTTGAAAGATTTTGTGTATTACTGA 1983
Db 4423 TATGGCTCCATTTTATTTTATAGTGTAAAGTTGATTTCCTAAAGTTTGTGTTTGTGCGA 4482

QY 1984 ACCTGTACTTTGTCTGA-----ATAGTTTATGGCAGCTATGATTCATGTTTAAAAA 2038
Db 4483 CAGTATCTTTTAAATGAGTCTTAAAAATAAAGGCATATTGTTTCATGTTTAAAAA 4542

QY 2039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2098
Db 4543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4602

QY 2099 AAAAAAAA 2106
Db 4603 AAAAAAAA 4610

```

Search completed: September 24, 2003, 05:56:49
Job time : 1401 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 06:06:25 ; Search time 418 Seconds
(without alignments)
3680.635 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSTEQMLRDPYRSMQING.....INTLKGGTDGKNTGHRNL 618

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 1678620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-DB=frame+p2n.model -DEV=xlh
-O=/cn2_1/USPTO_spool/US09856327/runat_23092003_153421_15588/app_query.fasta_1.775
-OB=PublishedApplications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09856327 @CN1_1.262 @runat_23092003_153421_15588
-NCPU=6 -ICPU=3 -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications_NA:

1: /cn2_6/ptodata/2/pubna/US07_PUBCOMB.seq:*
2: /cn2_6/ptodata/2/pubna/PCT_NEW_PUB.seq:*
3: /cn2_6/ptodata/2/pubna/US06_NEW_PUB.seq:*
4: /cn2_6/ptodata/2/pubna/US06_PUBCOMB.seq:*
5: /cn2_6/ptodata/2/pubna/US07_NEW_PUB.seq:*
6: /cn2_6/ptodata/2/pubna/PCTUS_PUBCOMB.seq:*
7: /cn2_6/ptodata/2/pubna/US08_NEW_PUB.seq:*
8: /cn2_6/ptodata/2/pubna/US08_PUBCOMB.seq:*
9: /cn2_6/ptodata/2/pubna/US09A_PUBCOMB.seq:*
10: /cn2_6/ptodata/2/pubna/US09B_PUBCOMB.seq:*
11: /cn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq:*
12: /cn2_6/ptodata/2/pubna/US09_NEW_PUB.seq:*
13: /cn2_6/ptodata/2/pubna/US10A_PUBCOMB.seq:*
14: /cn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq:*
15: /cn2_6/ptodata/2/pubna/US10_NEW_PUB.seq:*
16: /cn2_6/ptodata/2/pubna/US60_NEW_PUB.seq:*
17: /cn2_6/ptodata/2/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1091	33.2	1701	10	US-09-911-969-2 Sequence 2, Appli

2	177	5.4	1635	14	US-10-156-761-1918
3	177	5.4	9025608	14	US-10-156-761-1
4	148.5	4.5	1641	14	US-10-253-971-4
5	113	3.4	63158	12	US-10-292-198-1
6	112	3.4	1911	10	US-09-873-880-23
7	108.5	3.3	2346	14	US-10-353-445-8
8	108.5	3.3	3026	12	US-10-419-969-6
9	108	3.3	3431	14	US-10-128-714-75
10	108	3.3	3926	14	US-10-128-714-5075
11	107.5	3.3	9025608	14	US-10-156-761-1
12	107	3.3	4725	12	US-09-940-316B-24
13	106.5	3.2	3309400	10	US-09-738-626-1
14	106	3.2	3306	14	US-10-156-761-6845
15	106	3.2	68750	13	US-10-014-717-1
16	105.5	3.2	1542	14	US-10-156-761-1610
17	105	3.2	86114	15	US-10-080-170-648
18	104.5	3.2	1803	14	US-10-156-761-5001
19	104.5	3.2	32480	10	US-09-847-101B-23
20	104.5	3.2	32480	12	US-09-482-682-27
21	104.5	3.2	32798	12	US-10-424-638-1
22	104.5	3.2	34427	11	US-09-111-911-5
23	104.5	3.2	34555	12	US-10-117-982-479
24	104.5	3.2	35408	14	US-10-155-649-3
25	104.5	3.2	35871	9	US-09-956-335-2
26	104.5	3.2	35935	9	US-09-725-720-43
27	104.5	3.2	35935	10	US-09-782-378A-4
28	104.5	3.2	35935	10	US-09-782-378A-5
29	104.5	3.2	35935	12	US-09-739-007-43
30	104.5	3.2	35978	9	US-09-956-335-1
31	104.5	3.2	36620	11	US-09-952-060-30
32	104.5	3.2	37474	11	US-09-952-060-25
33	104.5	3.2	38519	11	US-09-952-060-28
34	104	3.2	4080	14	US-10-205-368-1
35	104	3.2	60196	14	US-10-205-032-1
36	103.5	3.2	1026	14	US-10-156-761-6467
37	103.5	3.2	36321	14	US-10-187-267A-1
38	103	3.1	5061	13	US-10-092-219-1
39	102.5	3.1	4611	10	US-09-801-368-321
40	102.5	3.1	4935	14	US-10-156-761-5059
41	102	3.1	3735	9	US-09-815-915-9
42	102	3.1	3735	12	US-10-393-316-9
43	102	3.1	5499	9	US-09-815-915-7
44	102	3.1	5499	12	US-10-393-316-7
45	102	3.1	26173	14	US-10-114-170-69

ALIGNMENTS

RESULT 1

US-09-911-969-2

; Sequence 2, Application US/09911969

; Patent No. US20020137896A1

; GENERAL INFORMATION:

; APPLICANT: Kawamura, Yukio; Morita,

; Akihiro; Izumo, Koji.; Saka, Tomohide.

; TITLE OF INVENTION: ANTITUMOR PROTEIN AND

; GENE ENCODING SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNIGAN, L.L.P.

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD 97

; CURRENT APPLICATION DATA: US/09/911,969

; APPLICATION NUMBER: US/09/911,969

; FILING DATE: 24-Jul-2001

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/023,731
;; FILING DATE: 13-FEB-1998
;; APPLICATION NUMBER: JP 29275/1997
;; FILING DATE: 13-FEB-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KENNETH H. SONNENFELD
;; REGISTRATION NUMBER: 33,285
;; REFERENCE/DOCKET NUMBER: 3479-4000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;;
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1701
;; TYPE: nucleic acid
;; TOPOLOGY: linear
;;
;; MOLECULE TYPE: (D) STRANDNESS: Single
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-911-969-2

Alignment Scores:
Pred. No.: 2,05e-110 Length: 1701
Score: 1091.00 Matches: 248
Percent Similarity: 57.04% Conservative: 84
Best Local Similarity: 42.61% Mismatches: 184
Query Match: 33.22% Indels: 66
DB: 10 Gaps: 15

US-09-856-327-2 (1-618) x US-09-911-969-2 (1-1701)

```
QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
|||||
DB 97 GATGTTTTCATTCTGGCAGTGGTCCCATAGTGTACTGTACCGCCGCACATCATGTAC 156
|||||
QY 57 -----AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAlaSerPheTyrAla 74
|||
DB 157 AATACCTCAACTACAAAGGTTTACATGGCGGAATAGGTCTCAAGATAAC-----207
|||||
QY 75 ValAsnAlaGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGlu 94
|||||
DB 208 -----CCTGTCATC-----GGGCCCATCACAGAATCTC 237
|||||
QY 95 IleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGln 114
|||||
DB 238 AATAAGTTTCAGAAAGACATTCACAAGTTTGTGAATATCATCAACGGTGCCCTCCAGCG 297
|||||
QY 115 ValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAla 134
|||||
DB 298 ATTTTCGATTTCGCCATCGGACACCTACACGCCCACTCTCGCTGTAGCAGCGTGG---GGG 354
|||||
QY 135 ProProGlySerSerAla-----IleSerAsnGlyLysAsnProHisGlnArg 150
|||||
DB 355 CCGCCCATCATCTCCGAGCCAGCCAGCTCGTGATTATGGACACCAATCCGAATCAGGAG 414
|||||
QY 151 GluPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyMetSerThrHis 170
|||||
DB 415 GCGGGCCTGAACCTTCCCGTAGCGTGTCACTAGGACAGATCGGGGGGATGGGACCCAC 474
|||||
QY 171 TrpThrCysSerThrProArgIleHisProProMetGluSerLeuProGlyIleGlyArg 190
|||||
DB 475 TGGACTTGGCGGTGTCTACTCCACATGAC-----504
|||||
QY 191 ProLysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGlu 210
|||||
DB 505 GAAGAGGGGTCAACACCCAGTGT---GACACAGGAGGTTCAGCCACTGCTCGAAGCT 561
|||||
QY 211 AlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeu 230
|||||
DB 562 GCTAAACATTCGTCACACGTTTACACGCGCAGTACGACGATTCATCCGTCAGATAGTT 621
|||||
```

```
QY 231 ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuPro 250
|||||
DB 622 GTCAAGAGAGACTTTCAGCAGAGACCCCT---GATGCGTCGCGGGTGTGACCACTCCCG 678
|||||
QY 251 LeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsn 270
|||||
DB 679 CTGGGGGTGAGCGCGGTACGGACAATCTATTTATGTACCTGACCGGTCGCGAT---735
|||||
QY 271 LeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuLeuThrAsn 290
|||||
DB 736 -----ACCGTCTTGGTGATGTGCCGAGAGAGTCCCGGATTCCTTGGTTACAGAG 786
|||||
QY 291 HisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGlu 310
|||||
DB 787 ACAGAGATGACGAAGCTTATTGTTCATGAAACCAATCCGACGAGGTGTGTGCGTTG 846
|||||
QY 311 ValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAla 330
|||||
DB 847 CTACGTTAACTTG-----AATCAAGCAAC-----GATGAATTTGTGCGGCC 888
|||||
QY 331 LysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGly 350
|||||
DB 889 AAGAGTTTCGTATAGCTTGTGGAGCAGTCTGCACACCGCAATCTTGTGGAACAGC---945
|||||
QY 351 PheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyr 370
|||||
DB 946 -----AACATCGCCCATAT-----CGCGTTGGTCGCTAC 975
|||||
QY 371 IleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSer 390
|||||
DB 976 CTCAGGAACAGTCCATGACTTTTGTGAGATCGTTCTCAAGAGGGGCATAGTGCATGCC 1035
|||||
QY 391 ValArgAspAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGlnHisIleAla 410
|||||
DB 1036 ATCGCTACTGACCTCGC-----TTGCTGCGAAGGTTGAGCGGCACAGAAG 1083
|||||
QY 411 LysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGlnValThrThr 430
|||||
DB 1084 AAGCACCCTGATGATGCTGTCCTATTCATTCACGAGGCTCAAGCTCAAGTGATGATT 1143
|||||
QY 431 ProPheThrGluGluHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGly 450
|||||
DB 1144 CCGTACACCTCGGACTTCCCTTGGCATGTTGAGTGCATCGCATCATCTCATATGTT 1203
|||||
QY 451 AlaValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThr 470
|||||
DB 1204 GATGTTGGACCAAGGCCGACCGCGCTGTTGTCGTGATGTAGGTTTTCGCAAAATCA 1263
|||||
QY 471 AspProGluAlaAsnAsnLeuLeuValPhe-----Gln 481
|||||
DB 1264 GATATTGTGGAAGAAATCGAGTGTACTTTGCTCGCAACCTTAAGTACCGGAGTGGGA 1323
|||||
QY 482 AsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThr 501
|||||
DB 1324 GCGGGTGTACAGACACTTATGGAATGCCACAGCCGACATTCATCTCAAGCGGACCAAC 1383
|||||
QY 502 AlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeu 521
|||||
DB 1384 GCGGTGGAGACCGTGCACAGAGGATGATGATGATATGATACCAACCTCGCAACATGCTG 1443
|||||
QY 522 GlyGlyTyrLeuProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHisLeu 541
|||||
DB 1444 GGTGGTACTCTCTCGCTCCCTACCTCAATTTATGGCACCTGGTCTCGTATGCAATC 1503
|||||
QY 542 AlaGlyThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeu 561
|||||
DB 1504 ACGGGAACCTACTCGGATCGGACAGATGATCAAACTTCTGTGCTGATCCGACATCAAAG 1563
|||||
QY 562 ValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGly 581
|||||
DB 1564 GTTCATAACTTCAACAATCTGTGGTTCGCGGGGATGGGTGATCCAGATCGGACTGCC 1623
|||||
QY 582 GluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsn 601
|||||
```

Db 1624 TCACACCGGACTCGTACGAGCGTCGCTATCGGCTCAAGGGTCTGAGGCTGTAGTCAAT 1683
Qy 602 ThrLeu 603
Db 1684 TACCTT 1689

RESULT 2

US-10-156-761-1918
; Sequence 1918, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1918
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1635)
US-10-156-761-1918

Alignment Scores:

Pred. No.: 2,54e-09 Length: 1635
Score: 177.00 Matches: 144
Percent Similarity: 32.60% Conservative: 63
Best Local Similarity: 22.68% Mismatches: 249
Query Match: 5.39% Indels: 179
DB: 14 Gaps: 26

US-09-856-327-2 (1-618) x US-10-156-761-1918 (1-1635)

Qy 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
Db 73 GAGGTCATCATATTGGCACCAGCGGTGGCGGCTGGCAGCGCTGCCACCGACTGGCCCCC 132
Qy 57 AlaGlyLeuArgValValMetValGluIleGly-----AlaAlaAspSer 71
Db 133 ACCGGGAACGGATCTCTGCTCGAACCGGCTGACTACCTGCCCGCGGAACGCGACAC 192
Qy 72 PheTyrAlaValAsnAlaGluGlyThrAlaValProTyrValProGlyTyrHisLys 91
Db 193 TGG-----GAATCCACCGCGGTCTTTCGTCGAAGGCAATACCGC 231
Qy 92 LysAsnGluIleGluPheGlnLysAspIleAspArgPhe-----ValAsnValIle 108
Db 232 GTCGCGAGTTCTTGTACGACACACACGGAACACAGTTCGCCCGCGCAATTTACTAC 291
Qy 109 LysGly-AlaLeuGlnValSerValProValArgAsnGlnAsnValProThrLeuAs 128
Db 292 GTCGCGGGAACACACCAAGTTCTACCGCGCGCGGCTCTTCGCTCGCGCCCGAGGACTTC 351
Qy 128 pProGlyAla-----TrpSerAlaProGlySerSerAlaIleSerAsnG1 144
Db 352 GCGGAACCTCGGCCACACGAGCGCATCTCCCGCGGCTTGCGCGCTGAGCTACGAGGACTC 411
Qy 144 yLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaValThrArgGlyVa 164

Db 412 GAGCGGTACTACACGAGCGCCGAACACCTCTACCTCG----- 448
Qy 164 lGlyGlyMetSerThrHisTrpThrCysSerThr---ProArgIleHisProProMetG1 183
Db 449 -----TCCACGGCGCGCACGCGGAAGACCGCGCGAGGTC---CCACGAGCGC 495
Qy 183 uSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAspLysG1 203
Db 496 CAGTACGGCT---ACCCGCGCGTCCAGCACGAGCGCGCATCCAGCACTACGACCGAC 552
Qy 203 uTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLysGluPhe-A 223
Db 553 CTGG-----AGAAAGCGGGCTCGACCCCTTCACCTTCGCTCCG 588
Qy 223 spLuSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArg 243
Db 589 ATCGGGGTCAACCTCACCA-----GGACGACCGGGCGGGCCACCCACGCC 636
Qy 243 lnaArgIlePhe-----ArgProLeuProLeu----- 251
Db 637 AGCGCTGCATCCGCTGGCGACCGGTGCGACGGCTTCCCTGTCTGTGTCGGCGCGAAGTCC 696
Qy 252 ----AlaCysHisArgLeuLysAsnAlaPro-GluTyrValGluTrpHisSerAlaGlu 269
Db 697 GACGCGCAGGTCTCTGTGTCGACCCCGCGGCGGAGCACGCC-----AACGTGCGAG 747
Qy 270 AsnLeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuLeuThr 289
Db 748 ATGCTCACCCATGCG-----GATGTGCGGGCGGCTCGACACGCGCGCACC 792
Qy 290 AsnHisArgCysThrArgLeuAlaLeuThrGlyTyrGluLysLysIleGlyAlaAla 309
Db 793 GGACGAGGTGCACCTCGCTCGCGACGGTGGG-----CACGGGCGCGCT 840
Qy 310 GluValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMet 329
Db 841 TCCACCGTGGAG-----TTCACG 858
Qy 330 AlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSer 349
Db 859 GCCGACATCGTGTGCTGCGCGCGCGTCACTCCGCGCTCTGTGCTGCGCTTCG 918
Qy 350 GlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArg 369
Db 919 GCC----- 921
Qy 370 TyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAsp 389
Db 921 ----- 921
Qy 390 SerValArgAspAsp-----ProTyrGlyLeuProTyrTrpLysGluAlaValAlaGln 407
Db 922 -----GATGACCGGCATCCGCGGCGCTGGCCACAGCTCGGGCGGTGGCGCGG 972
Qy 408 HisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGln 427
Db 973 CACTACATCGGCGCACAAACCTGCGCTGATGGCGGTGTCACCAAGGAACCGAACACACC 1032
Qy 428 ValThrThrProPheThrGluGluHisProTyrHis----- 439
Db 1033 AAATTCAGAGACCCCTGGCGCTGCACGACTGGTACCTGGGATCCGACGACTGGGACTAC 1092
Qy 440 -----ThrGlnIleHisArgAspAla 446
Db 1093 CCTCTCGGGCGCATCCAGATGCTCGGCAAGTCCGACGCGCGACGACATCCACGCGGCGG 1152
Qy 447 PheSerTyr---GlyAlaValGlyProGluValAspSerArgValIle----- 461
Db 1153 CCCCGCTGGCGCGCGCGCTCGCCCGACATCGCTTCGAGGTACTCGCCCCACCGACCGG 1212
Qy 462 ValAspLeuArgTyrPheGlyAlaThrAspProGluAlaAsnAsnLeuLeuValPheGln 481
Db 1213 GTCGACTTCTGGCTGTGGCGGAGAGACCTTGCCCCCTCGCGGAGAACCGCGCTCACCTGGAC 1272

QY 482 AsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThr 501
 Db 1273 GGGAGCGCGGCATCCATCTGGCGCTCGACGAG-----AAGAACAACATC 1317
 QY 502 AlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeu 521
 Db 1318 GCCGGGCTGAAGCGCTGGCGGCAACAACACTACAGGGCATG-----CTCAGCCACTTG 1368
 QY 522 GlyGly-----TyrLeuProThrSerProProGln 531
 Db 1369 GGCATGCACGAGCACCATCTGCTGTCGCACAGCATCTACCTGCACAAGGGCATGCC--- 1425
 QY 532 PheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLys 551
 Db 1426 -----ATCGCGCCACCGCGCATCGAGCGGCGAGGTTCGCTTCGCGCGCGACCCC 1476
 QY 552 AlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla 571
 Db 1477 GCGAGCTCCCGCTCGACGTCACCTGCAAGGCCACGACCTCGACAACTCTACGTCGTC 1536
 QY 572 GlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
 Db 1537 GACACGAGCTTCTCCCGAGCATCGGCGGCTGAATCCCTCGCTGACCGCGCATCGCCAAC 1596
 QY 592 AlaIleLysSerAlaArgSerIleIleAsnThrLeuLys 604
 Db 1597 GCCTCTCGGGTCGGCGACCATCTCGCGGAGGACTGCGG 1635

RESULT 3
 US-10-156-761-1
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156, 761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Alignment Scores:
 Pred. No.: 0.00104 Length: 9025608
 Score: 177.00 Matches: 144
 Percent Similarity: 32.60% Conservative: 63
 Best Local Similarity: 22.68% Mismatches: 249
 Query Match: 5.39% Indels: 179
 DB: 14 Gaps: 26

US-09-856-327-2 (1-618) x US-10-156-761-1 (1-9025608)

QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
 Db 2359134 GAGGTATCATCATTTGGACCGGTGGCGGGTGGCACGCTGCCCCACCGACTGGCCCCC 2359193

QY 57 AlaGlyLeuArgValValMetValGluIleGly-----AlaAlaAspSer 71
 Db 2359194 ACCGGGAACGGATCTCTGCTCGAAGCGGTGACTACCTGCCCGCGGAACCGACAAC 2359253
 QY 72 PheTyrAlaValAsnAlaGluGlyThrAlaValProTyrValProGlyTyrHisLys 91
 Db 2359254 TGG-----GAATCCACCGCGGTG---TTCGTCAAGGGCAATACCGC 2359292
 QY 92 LysAsnGluIleGluPheGlnLysAspIleAspArgPhe-----ValAsnValIle 108
 Db 2359293 GCTCCGGAGTTCGTGTACGACAACACGGAACAGTTCCTCCGCCCGGAAGTCAATACTAC 2359352
 QY 109 LysGly-AlaLeuGlnValSerValProValArgAsnGlnAsnValProThrLeuAs 128
 Db 2359353 GTCGCGGCAACCAAGTCTACGCGCGCGCTCTTCGCTGTCGCGCGCGAGGACTTC 2359412
 QY 128 PProGlyAla-----TrpSerAlaProProGlySerSerAlaIleSerAsnG 144
 Db 2359413 GCGCAACTCCGCCACACGACGCGCATCTCCCGCGCTGCGCGCTGAGCTACGAGGAATC 2359472
 QY 144 YLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaValThrArgGlyVa 164
 Db 2359473 GACCCGTACTACGACGCGCGCAACCTCTACTCTG----- 2359509
 QY 164 lGlyGlyMetSerThrHisTrpThrCysSerThr---ProArgIleHisProMetG 183
 Db 2359510 -----TCCACGCGCGGCACGCGAAGACCGCAGCGAGGTC---CCACCGCGCC 2359556
 QY 183 uSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAspLysG 203
 Db 2359557 CAGTACGCGCT---ACCGCGCGGTCCAGCAGCAGCGCGCGCATCCAGCAACTCAGCCACGAC 2359613
 QY 203 vTrpAsnGluLeuTyrSerGluAlaGluArgIleGlyThrSerThrLysGluPhe-A 223
 Db 2359614 CTGG-----AGAACAGCGGTGTCGCCCTTCCACCTTCCG 2359649
 QY 223 spGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArg 243
 Db 2359650 ATCGGGTGAACCTCACCCA-----GGACGACGCGGCGCGGCCACCCACGCC 2359697
 QY 243 lNArgIlePhe-----ArgProLeuProLeu----- 251
 Db 2359698 AGCGCTGCATCCGCTCGCAGCCGCGTCCCGCTTCCCTGCTGTCGCGCGGAAGTCC 2359757
 QY 252 -----AlaCysHisArgLeuLysAsnAlaPro-GluTyrValGluTrpHisSerAlaGlu 269
 Db 2359758 GACGCGCAGGTATCTGTGTGACCCCGCGCTGAGCAGCGCC-----AACGTCGAG 2359808
 QY 270 AsnLeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuLeuThr 289
 Db 2359809 ATGCTCACCATCGC-----GATGTGCGCGCTCGACACGCGCGGACCC 2359853
 QY 290 AsnHisArgCysThrArgLeuAlaLeuThrGlyTyrGluLysLysIleGlyAlaAla 309
 Db 2359854 GGACGGAGTGTCACTCGGTGCGCGCGGTGGG-----GACGCGCGCGCT 2359901
 QY 310 GluValArgAsnLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMet 329
 Db 2359902 TTCACCGTGGAG-----TTCAGC 2359919
 QY 330 AlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnLeuTyrAsnSer 349
 Db 2359920 GCGACATCGTGTGTCGCTCGCGCGCGCTCACTCCCGCTGCTGCTGCTGCTGCTGCTGCTG 2359979
 QY 350 GlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArg 369
 Db 2359980 GCC----- 2359982
 QY 370 TyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAsp 389
 Db 2359982 ----- 2359982

```
Qy 390 SerValArgAspAsp-----ProTyrGlyLeuProThrTrpTrpLysGluAlaValAlaGln 407
Db 2359983 -----GATGACGGCATCCGCGAGCGCTGGCAACAGCTCGGGCGTGGTGGCGG 2360033
Qy 408 HisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGln 427
Db 2360034 CACTACATGCGGCAACAACCTGGCCCTGATGGCCGTGTCGAAGGAACCGAACGACACC 2360093
Qy 428 ValThrThrProPheThrGluGluHisProThrHis----- 439
Db 2360094 AAATCCAGAGACCCTGGCCCTGCACAGCTGGTACCTGGGATCCGACGACTGGGACTAC 2360153
Qy 440 -----ThrGlnIleHisAspAla 446
Db 2360154 CCTCGGCGGCATCCAGATCTCGGCAAGTCCGACGCGGAGATCCACGCGGAGGCG 2360213
Qy 447 PheSerTyr---GlyAlaValGlyProGluValAspSerArgValIle----- 461
Db 2360214 CCGCGTGGCGGGCGCGTCCGCCGACATCGCTGGAGGTACTCGCCACACACGCG 2360273
Qy 462 ValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGln 481
Db 2360274 GTCGACTTCTGGCTGCGGAGAGGACCTGCCCTCGCGGAGAACCGCGTCACTCGGAC 2360333
Qy 482 AsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThr 501
Db 2360334 GGGGACGCGGCATCCATCTGGCGCTGCAGCAG-----AAGAACAACATC 2360378
Qy 502 AlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeu 521
Db 2360379 GCGGGGTGAAGCGCTGCGGCACAACTACAGGGCATG-----CTCAGCCACTTG 2360429
Qy 522 GlyGly-----TyrLeuProThrSerProGln 531
Db 2360430 GGCATGACGAGCACCATCTGCTGTCGACAGCATCTACCTGCACAGGGCATGCC--- 2360486
Qy 532 PheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLys 551
Db 2360487 -----ATCGGCGCACCGCGCATCAGCGGGCAGCGTCCGCTTCGCGCGGACCCC 2360537
Qy 552 AlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla 571
Db 2360538 GCCACTCGCGCCCTGAGTCACTGCAAGGCCACGACCTCGACCACTCTAGTCGTCTC 2360597
Qy 572 GlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
Db 2360598 GACACGAGCTTCTTCCCGAGCATCGCGCGGTGATCCCTCGCTGACCGCATCGCCAAC 2360657
Qy 592 AlaIleLysSerAlaArgSerIleAsnThrLeuLys 604
Db 2360658 GCCCTGGGGTCGCGCACCATCTCGCGGAGCGACTGGGG 2360696

RESULT 4
US-10-253-971-4
; Sequence 4, Application US/10253971
; Publication No. US20030070192A1
; GENERAL INFORMATION:
; APPLICANT: GEORGES, FAWZY
; APPLICANT: DONG, JIN-ZHUO
; APPLICANT: KELLER, WILF
; APPLICANT: HUSSAIN, ATTA A. K.
; APPLICANT: SELVARAJ, GOPALAN
; APPLICANT: DATLA, RAJU
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING SECONDARY PLANT
; FILE REFERENCE: 73776-159D
; CURRENT APPLICATION NUMBER: US/10/253,971
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: US 60/072156
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: US 09/012453
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 7
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1641
; TYPE: DNA
; ORGANISM: Arthrobacter pascens
US-10-253-971-4

Alignment Scores:
Pred. No.: 3.63e-06 Length: 1641
Score: 148.50 Matches: 128
Percent Similarity: 32.90% Conservative: 75
Best Local Similarity: 20.75% Mismatches: 238
Query Match: 4.52% Indels: 177
DB: 14 Gaps: 27

US-09-856-327-2 (1-618) x US-10-253-971-4 (1-1641)
Qy 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyr---AlaLysLeuCysVal 55
Db 43 GACTACATCATCGCGCGGTTCGCCGAGCGGCGAGTCCGCCGCTGAGCGAG 102
Qy 56 GluAlaGlyLeuArgValMetValGluIleGlyAlaAlaAspSerPheTyrAlaVal 75
Db 103 GAGCCACCGGTTCGTGGCGCTGGTGGAGCGCGCGGACGAC----- 147
Qy 76 AsnAlaGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIle 95
Db 148 -----CGCGCGTTCGCGAGGTA----- 165
Qy 96 GluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnVal 115
Db 166 -----CTGAGCTGACGCGTGGTGGAGTGTGTAATCCGCTAGCAG----- 210
Qy 116 SerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaPro 135
Db 211 -----TGGACTATCCCG 222
Qy 136 -----ProGlySer-SerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGln 153
Db 223 ATCGAACCGCAGGAGAACGCGCACTCTTCATCGCCCGCCCGCGCGAGATCATGGGT 282
Qy 153 uAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrCy 173
Db 283 GGCTGCTCCACCACTCTCGCATCGCTTCCTGGCGCGCGCGAGAGACTGGACGAG 342
Qy 173 sSerThrProArgIleHisProMetGluSerLeuProGlyIleGlyArgProLysLe 193
Db 343 TGGGAGTCCAAAGTACGGCGCGCCGCTGGAGAACGCTG----- 379
Qy 193 uSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluAr 213
Db 380 ---AGTCCGCTGCGCGCTTACCAGCGGCTGGAGACCAACGAGGACGCGCGCGGAC 435
Qy 213 gLeuIleGlyThrSerThrLys----- 220
Db 436 GCGCGCACCGCAGCGGCTGACGCGCGCTGACCTGATGAACGCTGCCCGCGCGACCCC 495
Qy 221 -----GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGln 236
Db 496 GCGCGGTGCGACTCTCTGGAGCGCTGCAACAGCAGGACGATTCGCGCGCGGAGTTCAAC 555
Qy 236 nAspAla-TyrLysAspArgGlnArgIlePheArgPro-----LeuProLeuAlaCysH 254
Db 556 ACCGCGACCGCGTCAATGCGCCCACTTTTCCAGATCACACGCGCGCGGACGCG 615
Qy 254 is-ArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHis 273
Db 616 ACCCGTTCCTCCAGCTCGGTCTCTACATC-----CAG 648
Qy 274 SerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCys 293
Db 649 CCGATCATC-----GAGCGGGGAACCTTCACTCCCTGCTGACCGGGTTCGCGGCC 696
```

```

QY 294 ThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsn 313
Db : : : : :
697 CGGCAACTGCTTGCAGCG-----GACAAGCGTGCACCGCGTGCACCTTGTGGAC 750
QY 314 LeuLeu-----AlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLys 331
Db : : : : :
751 TCGGGGTTCGGCGGACATACCGGCTCTCCGGCGCTTGGAG----- 792
QY 332 ValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPhe 351
Db : : : : :
793 ---GTCATCTGTCCACCGGCCCATGTACTCGCTAGCTGCTATGCTCTCCGGCATC 849
QY 352 Ser-----GlyLeuGlnValThrProArgAsnAspSerLeu 363
Db : : : : :
850 GGCCCGCGCGCACTCGCCGAGCAGCGGCTCGAGGTCTGTGTC-----GACTCC--- 900
QY 364 IleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeu 383
Db : : : : :
901 ---CCGGTGTCCGGAGCAGCTGCAGGACACCCGGAAGCGTGTCTCAGTTCGAGGCC 957
QY 384 ArgGlnGluPheValAspSerValArgAspProTyrGlyLeuProTyrTrpLysGlu 403
Db : : : : :
958 AAGCAGCAGATGTCAGACT-----TCGACCGAGTGTGGGA-GAT 998
QY 404 AlaValAlaGlnHisIle-----AlaLysAsnProThrAsp 415
Db : : : : :
999 -----CGCATCTTCACCCCCACCGAGAACGCGCTGGACCGCGCGGACCTGATGAT 1049
QY 416 AlaLeuProIle---ProPheArgAspProGluProGlnVal----- 428
Db : : : : :
1050 GCATCAGCGCTCGCTCGCTGCATGAACACCTCGCGTACGGTACCCCAACCCAGCA 1109
QY 429 -----ThrThrProPheThrGluHisProTyrHisThrGlnIleHisArgAspAla 446
Db : : : : :
1110 GAACGGCTTCACCTCAGCCGAAAGCTCAGCAGCCCGCTCCCGGGCAGCTCGC--- 1166
QY 447 PheSerTyrGlyAlaValGlyProGluValAsp-SerArgValIleValAspLeuArgTr 466
Db : : : : :
1167 -----GCTCGCGAGCGCGAGCTTCGCGACAAAGCGCGCGTTCGACCCGCGGTA 1214
QY 466 pPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAs 486
Db : : : : :
1215 CTTT-----ACTGATCGGAGGCGCAGCATGCGCGTATGGTGGCGCGCATCCGC--- 1266
QY 486 pGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsnValAr 506
Db : : : : :
1267 -----AAGGCCGCTGTAATCGC 1283
QY 506 gAlaArgLysMetMetAlaAspMet-----CysG1 516
Db : : : : :
1284 CCGCCAGCGCTGCATGCGCGAATGAGCGCGCGGAGCTCTCGCCCGCAGCGAGCGCA 1343
QY 516 uValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProGlnPheMetAspProG1 536
Db : : : : :
1344 GACCGACAGGAACTGCAGGACTACATCCGCAAGACG-----CACAA 1385
QY 536 yLeuAlaLeuHisLeuAlaGlyThrArgIleGly---PheAspLysAlaThrVa 555
Db : : : : :
1386 CACCGTTTACCACCCCGTGGCAGCGTCCGATGGACCGACCGCAGCATGTCGCC 1445
QY 555 lAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyTh 575
Db : : : : :
1446 GCTCAGCCCGAGCTGCGGGTGAAGGGGTGACCGCGCTGCGCGTATGCTCTGT 1505
QY 575 rIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMet 589
Db : : : : :
1506 CATGCTGACACAGTCACTCGGTCATCCCAACATCACCTCATG 1548

```

RESULT 5

US-10-292-198-1/c
; Sequence 1, Application US/10292198
; Publication No. US2003015765A1
; GENERAL INFORMATION:

```

; APPLICANT: SHEN, Ben
; APPLICANT: LIU, Wen
; TITLE OF INVENTION: BIOSYNTHESES OF ENEDIYNE COMPOUNDS BY MANIPULATION OF C-1027 GE
; FILE OF INVENTION: PATHWAY
; FILE REFERENCE: 054030-0007
; CURRENT APPLICATION NUMBER: US/10/292,198
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 10/159,257
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 09/478,188
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: US 60/115,434
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 63158
; TYPE: DNA
; ORGANISM: Streptomyces globisporus
US-10-292-198-1

```

```

Alignment Scores:
Pred. No.: 7,34 Length: 63158
Score: 113.00 Matches: 136
Percent Similarity: 33.8% Conservative: 78
Best Local Similarity: 21.52% Mismatches: 249
Query Match: 3.44% Indels: 171
DB: 12 Gaps: 29

US-09-856-327-2 (1-618) x US-10-292-198-1 (1-63158)

QY 84 ProTyrValProGlyTyrHisLysLysAsnGluIleGluPheGlnLysAspIleAspArg 103
Db : : : : :
60129 CGCGCTCCACCCGACACACGCTCCACCGGGAGCCCGCATGATGAGCAGCAGCCGCGGCGC 60070
QY 104 Phe-----ValAsnValIle-----LysGlyAlaLeuGlnGlnValSerValPro 118
Db : : : : :
60069 ACGGACGCTCTGCTCATCGGACGCGATTCGGCGGTGCATCGCGCGCTACACCTGCGC 60010
QY 119 ValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProPro-Glyse 138
Db : : : : :
60009 CGGGGGCGCGCAGCATCCGCTCTGGAACCGCGCGCTGCTGAGAGCAGGAGTTC 59950
QY 138 rSerAlaIleSerAsn-----GlyLysAsnProHisGlnArgGluPh 152
Db : : : : :
59949 GAGCAGCACTACAAACTCGGCTCCTCGTACAGAGGCGGTTCCGACTTCAGGTCGGGAC 59890
QY 152 eGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpTh 172
Db : : : : :
59899 GGCATGAGCATCTCGCGGGGAACTGTGTGGCGCGCGCAGCGTCTACTTCGCCGCG 59830
QY 172 rCysSerThrProArgIleHisProProMetGluSerLeuPro-----GlyIleG1 189
Db : : : : :
59829 ATGCCCGCGCGCGCGCTTCTCTTCAGCCGCGAGGGTCCATAGGACCGCGGATGTGG 59770
QY 189 yArgProLysLeuSerAsnAspProAlaGluAspLysGluTrpAsnGluLeuTyrse 209
Db : : : : :
59769 CGCGAGCGGTGAGCGCGGAAACCC-----TGGACCCCTGG-TACGA 59729
QY 209 rGluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGlu--SerIleArgHis 228
Db : : : : :
59728 CCGTGTCTGAGGAGTCCCTCTCCGTACCGCGGACGACTGGAAACGACGTGAGTACGCGCG 59669
QY 229 ThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgPro 248
Db : : : : :
59668 CGGACTGTGGCGCGCGGTGCAACCA-----CGCGCGCGCGCAGCCCA 59624
QY 249 LeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGlu-TrpHisSerAl 268
Db : : : : :
59623 CCCCTCTCGCTCGGCATCGACAAACCAAGTGCCTGAA---CTGCAACTGGATGATGCG 59567
QY 268 aGluAsnLeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuLe 288
Db : : : : :

```

Db 59566 CGGCTGCCGCTTCGAGGCC-----AAGCAG-----TCCTTGCT 59534
QY 288 uThrAsnHisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAl 308
| | | | | : : : : :
Db 59533 GGTCAACTACCTGCCCGCGCATCGCC-----CATGGGCG 59498
QY 308 aAla-----GluValArgAsnLeuLeuAlaThrArgAsnProSerSerG1 323
| | | | | : : : : :
Db 59497 CGGCATCGCGCGCTGTCAGAGGTCCAGCACCTCTCGCGCACCCCGCGAGCTCTACCG 59438
QY 323 nLeu-----AspSerTyr-----II 328
| | | | | : : : : :
Db 59437 GGTGCATACAACTGTCGACGACGAGTACTCCGCTCCAGCGCGGCGAGCGGTCT 59378
QY 328 eMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAs 348
| | | | | : : : : :
Db 59377 CGAGCGGAAGATCGTGGTGATGGCGCGCGCGCGCGCGGTCCTCCAGCG 59318
QY 348 nSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuG1 368
| | | | | : : : : :
Db 59317 CAGCGAG---GCCACCTGGGCACCATGCCAGA-----GCGGTGGG 59279
QY 368 yArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheVa 388
| | | | | : : : : :
Db 59278 ACGTACTTCTCCGGCAACGCGGAGCGG-----CTCAACACCGCATCAT 59234
QY 388 lAspSerValArg-----AspAspProTyrGlyLeuProTyrTrp 401
| | | | | : : : : :
Db 59233 CGACGAGCCAGGCGCGGAGCTGTCGGCTGACCGCGGCGGCGGCTCGCTAC-- 59176
QY 401 pLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAla-LeuProIlePro 421
| | | | | : : : : :
Db 59175 -----GGGGCAACCATGATCGCAAGGGGCCCGTGGCGAGCTGGACAGGCT 59126
QY 421 heArgAspProGluProGlnValThrThrProPheThrGluHisProThrHisThrG 441
| | | | | : : : : :
Db 59125 GGACGGCTGCTGGCCCG-----AGTACTCCGCTACT-----CCCTGGAACAGCT 59081
QY 441 nIleHisArgAspAlaPheSerTyr-----GlyAlaValGlyProG 455
| | | | | : : : : :
Db 59080 CTACTTCCGCGCGGCTCGGCACG-ATCCTCGCCAGGTGCCGCGCCAGCGGCCCT 59022
QY 455 luValAspSerArgValIleValAspLeuArgTyrPheGlyAlaThrAspProGlu---- 473
| | | | | : : : : :
Db 59021 CC-----TGGTTCGGCAAGGAGAGAGAGATCC 58992
QY 474 -----AlaAsnAsnLeuLeuValPheGln-----AsnAspValGlnAspGlyT 488
| | | | | : : : : :
Db 58991 TGAACGAGTGGACCTCTCGGCTCAGGATCTTCACGATGATCGAGGACGACGAAGCG 58932
QY 488 yrSerMetProGlnPro----- 493
Db 58931 TCTTCGGCCCTCCCGCCACCGGTACGCCACCGCATCTCCAGCAGATGCTCGGCC 58872
QY 494 -----ThrPheArgTyrArgProSer-----ThrAlaSerAsnV 505
| | | | | : : : : :
Db 58871 GGGGCAACCTCGCTTACGACCGAGCAACACCCCTGGCGCCCTGGCGCATCCGACG 58812
QY 505 alArgAlaArgLysMet-----MetAlaAspMetCysGluValAlaSerA 520
| | | | | : : : : :
Db 58811 CCGAGGTCAAGAGATCTCTGGAGAGGACGCGCTGGCCCAAGGTCTACCTCGGACCAACG 58752
QY 520 snLeuGlyGlyTyrLeuProThrSerProGlnPheMetAspProGlyLeuAlaLeuH 540
| | | | | : : : : :
Db 58751 ACCTCGTCGGC-----GCCACACCGTCC 58728
QY 540 isLeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsn 560
| | | | | : : : : :
Db 58727 ACCCGCTCTCTCTCGCGCGATGGCGAGACCCACACATCCCGCTCGACGACACCA 58668
QY 560 erLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyP 580
| | | | | : : : : :
Db 58667 ACGAGCTCCGGGACCAACCGCGGATCTTCGTCACCGAGGATCTCGGTCGCTCCCGCGGCG 58608

RESULT 6

US-09-873-880-23

; Sequence 23, Application US/09873880

; Patent No. US20020123118A1

; GENERAL INFORMATION:

; APPLICANT: Sewalt, Vincent

; APPLICANT: Falco, S. Carl

; APPLICANT: Allen, Stephen M.

; TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES

; FILE REFERENCE: BB192 US CIP

; CURRENT APPLICATION NUMBER: US/09/873,880

; PRIOR FILING DATE: 2001-06-04

; PRIOR APPLICATION NUMBER: 09/363,321

; PRIOR FILING DATE: July 28, 1999

; PRIOR APPLICATION NUMBER: 60/094,839

; PRIOR FILING DATE: July 31, 1998

; NUMBER OF SEQ ID NOS: 42

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 23

; LENGTH: 1911

; TYPE: DNA

; ORGANISM: Zea mays

US-09-873-880-23

Alignment Scores:

Pred. No.:	Score:	Length:
0.0498	112.00	1911
Percent Similarity:	32.50%	Matches: 132
Best local Similarity:	19.50%	Conservative: 88
Query Match:	3.41%	Mismatches: 214
DB:	10	Indels: 244
		Gaps: 33

US-09-856-327-2 (1-618) x US-09-873-880-23 (1-1911)

QY	2	SerLeuSerThrGluGlnMetLeuArgAspTyrProArgSerMetGlnIleAsnGlyGln	21
Db	84	TCGCTCAACTCTCCACATGACTACTAGTAT-----	116
QY	22	IleProLysAsnAlaIleHiscluThrTyrGlyAsnAspGlyValAspValPheIleAla	41
Db	117	CTTCCCGCTCTGCC-----AGTCGCGCTAGCTATATCATCGTA	158
QY	42	GlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAlaGly	58
Db	159	GGTGGTGGCAGCGCTGGA-----TGTGTCTGGCTTCCGCCCTATCC	200
QY	59	-----LeuArgValValMetValGluIleGlyAlaAspSerPheTyr	73
Db	201	TCCTACCTTCCTGAGCGCAAGGTTTATGATTGAGCTGGCCCTTCAGAC-----TTC	254
QY	74	AlaValAsnAlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsn	93
Db	255	GGTCTCAAC-----AAT	266
QY	94	GluIleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGln	113
Db	267	GTCCTGAACCTTCGGGAG-----TGGCTGTCTCTCTTGGTGGTGATCTC-----	311
QY	114	GlnValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSer	133

1945	TTC-	-----ACAAACACCACCGCTTGCTC	1968
QY	495	eArgTyrArgProSerThrAlaSerAsnVal-----ArgAlaArgLysMe	510
Db	1969	ATCCAGTACGAGAACTTACCGCGACTGGATTGTCAACCACAACGTCGCGTACTCGGAAC	2028
QY	510	tMetAlaAspMetCysGlu---ValAlaSerAsnLeuGlyGlyTyrlLeuPro-----	526
Db	2029	TTCTCTGGACACTCCCGAGTAGCAGCTTCGATGTGTGGACCTTCTGCCCTTCAACCGA	2088
QY	527	-----ThrSerProGlnPheMetAspProGlyLeuAlaLeuHi	540
Db	2089	GGATAGTTCACATCTCGACAGGACCCCTACTTCACCACTTCGCCTACGACCCCTCAG	2148
QY	540	sLeuAlaGlyThr-----ThrArgIleGlyPheAspLysAlaThrThrValAlaAspAs	558
Db	2149	TACTTCTCAACGAGCTGGACCTG-CTCGGTGAGGTGCGCGTACTCAACTGGCCGCCAA	2207
QY	558	nAsnSerLeuValTrpAspPheAlaAsnLeuTyrrValAlaGlyAsnGlyThrIleArgTh	578
Db	2208	C-----ATCTCCAACCTCCGTCCTGCATGCAGAC	2234
QY	578	rGlyPhe-----GlyGluAsn	583
Db	2235	CTACTTCGCTGGGAGACTATCCCGGTGATAAC	2268

```

RESULT 9
US-10-128-714-75/c
; Sequence 75, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
; LENGTH: 3431
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-75

```

Alignment Scores:	
Pred. No.:	0.332
Score:	108.00
Length:	3431
Percent Similarity:	129
Best Local Similarity:	33.44%
Query Match:	Conservative: 84
DB:	Mismatches: 211
	Indels: 213
	Gaps: 36
	DB: 14

US-09-856-327-2 (1-618) x US-10-128-714-75 (1-3431)

Qy	37	AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAla	51
Db	2701	AAACACCTGGCTACCGGCTCCTCGATATATGGGTGTACGTATGCTTTAATCCTGGTAGC	2642

```
QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
Db 1717 -----AAGTCGGCGGTGGTGGACGCGCAAAATGTCTGCGGTACCTTG 1664
QY 341 -----GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThr 357
Db 1663 CGAGTGGCACAAAGCCCGAGCTCTCAA-----CCCGGTCTCAAACAGGAC 1616
QY 358 ProArgAsnAspSerLeuIleProAsnLeuGlyArg-----TyrIleThrGlu 373
Db 1615 CCGCGAAAC-----AAGCGCGTCAAGTCCAGCAGCCCATCTGCGCGAT 1571
QY 374 GlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAsp 393
Db 1570 GTTCCACGTCACATCCGCGGATCACAGTCAGTCGAGCGTTGTATG-----CTCCGCGT 1517
QY 394 AspProTyrGlyLeu-----ProTrpTrpLysGluAlaValAla 406
Db 1516 CATCCCGCAGCGATTGTGCTTAACAGAGCGTCTACCCGGCTTTGCAGAGATAGAAGC 1457
QY 407 Gln-----HisIleAlaLysAsnProThrAspAlaLeuProIleProPhe 421
Db 1456 CAGGCTCTGCTGCTTCACACAGAACGACCCAGCACTCTGCAACAGCCCATGAACAG 1397
QY 422 ArgAspProGluProGlnValThrPropheThrGluGluHisProTrpHisThrGln 441
Db 1396 CAGAACAGGATACCTCTGTTTCGACCCGAGGCAATTCGGCCTTCTCCGCCCAACAC 1337
QY 442 IleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluVal-----AspSer 458
Db 1336 GTCCGG-----GCCGGGGCCCTCTCTCTCTTCCGACAAAGG 1298
QY 459 ArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeu 478
Db 1297 CGGTACTCTCGGGGTAA-----AACATGCCAACAGTAAATATAAGCC 1250
QY 479 ValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnPro-----493
Db 1249 ATCCTCCGTGGAGCTTGAACGCTCTCGACATCGAGCCCTACCCGCGCGGTAAATACCT 1190
QY 494 -----ThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMet 510
Db 1189 AACATCACAAACGAGCAAGTCGGCGCCGCTCCCAAGCGGGACACTATCCGGCGCCTC 1130
QY 511 MetAlaAspMetCysGluValAlaSerAsnLeuGlyTyrLeuProThrSerProPro 530
Db 1129 CTGCTC-----CACATCCCGCTTCGCGCTCGCGCGCTTCCACCT 1091
QY 531 Gln-----PheMetAspProGlyLeuAlaLeuHisLeuAlaGly 543
Db 1090 TCTAAGCTCTCTCTCGACTCCACTGCTCGCTCGCGCTCTCTCGG 1046
RESULT 10
US-10-128-714-5075/C
; Sequence 5075, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Sebastian M
; APPLICANT: Lemieux,
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT FILING DATE: 2002-04-23
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR APPLICATION NUMBER: US 60/295,890
```

```
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5075
; LENGTH: 3926
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-5075

Alignment Scores:
Pred. No.: 0.406 Length: 3926
Score: 108.00 Matches: 129
Percent Similarity: 33.44% Conservatives: 84
Best Local Similarity: 20.25% Mismatches: 211
Query Match: 3.29% Indels: 213
DB: 14 Gaps: 36

US-09-856-327-2 (1-618) x US-10-128-714-5075 (1-3926)

QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAla-----51
Db 3196 AATACCTGGCTACCGGCTCTCGATATATATGGGTGTACGTATGCTTTTAACTCCTGGTAGC 3137
QY 52 -----LysLeuCys 54
Db 3136 AAGCGCAGGTAGTCTGGTCTGTGGTCGTGGTAGCAGCCGTGTGATTCCATCTCTGT 3077
QY 55 ValGluAlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPhe-----72
Db 3076 GTGGGCTTGGTATCGAC---GGAATGATGAGAGCTGGGATAGCTATAAGTCTCTGAATT 3020
QY 73 ---TyrAlaValAsnAlaGlu-----GluGlyThrAlaValPro---TyrValProG 88
Db 3019 AGATACGCTGTACATATGCAATGTCTTAATGACATCAAAGCCTGCCAGAACAGTCATAT 2960
QY 88 lYTrpHisLysLysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsnValI 108
Db 2959 CTCATCACCCTAGTCA-----TTCTGTGAAGGAC-----CGTGTCA 2924
QY 108 leLysGlyAlaLeuGlnGlnValSerVal-----117
Db 2923 TGAACACACCTTTTGAACCCGCGCAGCACTTCCGCGCGCTCGCATGGTCTGTATAT 2864
QY 118 -----ProValArgAsnGlnA 123
Db 2863 AACATCGCGGACCTCTTTGACCCCACTTCTCCACCGCATCCATGGCCACAGCATCTAG 2804
QY 123 snValProThrLeuAspProGlyAlaTrpSerAlaProProGlySerSerAlaIle----141
Db 2803 ATGCTCATACTCTTCAA---TCACTTTGAATGACACCACCTGCACATGGGCTCGGTCC 2747
QY 142 -----SerAsnGlyLysAsnProHisGlnArgGluPheGluAsnL 155
Db 2746 GCTGTGCAACCGCTGGAGGAGACGCGCGCTCGACCA-----GCGCATC 2702
QY 155 euSerAlaGlu-----AlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrC 173
Db 2701 TGACCCCGCAACCCACAGCAGCAAAAGGGGGCTCTCGCGCCCGCTACAGCGCGGTGTCAGT 2642
QY 173 ysSerThrProArg-----IleHisProPro-----181
Db 2641 GCGCGAGCCGCTGCTCTCGCTCGCGCTGCAGACGAGCATCTCTCTCCATCTCGCGGAG 2582
QY 182 -----MetGluSerLeuProGly---IleGlyArgProLysLeuSerAsnAspProAla-198
Db 2581 GGTCTCTTCCATCGTGGCCAGGATGACCGCTCGCTCGCGAAGCAGCAGCCCTGACCCAGCA 2522
QY 199 GluAspAspLys-----GluTrpAsnGluLeuTyrSerGluAlaGluArg 213
Db 199 GluAspAspLys-----GluTrpAsnGluLeuTyrSerGluAlaGluArg 213
```

Db 2521 CCACCGATCGTCTCGCTGCTGATACATACATGGCGCGAATTG-----GAACAT 2474
QY 214 LeuileGlyThrSerThrLysgluPheAsp-----GluserIleAargHisThr 229
Db 2473 GCGATCGCGGAGTCGCGCATCCACGCGACGTCGCTCCAGCGGAACAGGTGCGCGAAGAC 2414
QY 230 LeuVal-----LeuArgSerLeuGlnAspAlaTyrIlysAsp 241
Db 2413 ATGATACCCCAACGCCCGTAGTCGGGGATGGAGAAAGCGGTGGAGCGTTATCATCAT 2354
QY 242 ArgGlnArgIlePheArgProLeuAlaCysHisArgLeuLysAsnAlaProGlu 261
Db 2353 GGGGATGAA-----TCGGTGATGCCGAAATCGTCCGGAA 2318
QY 262 TyrVal---GluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspIlys 280
Db 2317 TATGTCAGCGGTATACTCGCATCAACCGAAGTACACCTCTC-----GATCAA 2267
QY 281 GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly 300
Db 2266 CGGCCC----- 2261
QY 301 GlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 320
Db 2260 GGCATACACGCCCGGTGCGAGGCGACAGAATACAGATATCTCTCGCC----- 2213
QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
Db 2212 -----AGTTCCGGCGGTGGTCTGGACAGCGCCACAATGTCTCGCGGTACCTTG 2159
QY 341 -----GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThr 357
Db 2158 CGAGTGGCACACAAGCCCGCTCTCAA-----CCGGTCTCAACAGGAC 2111
QY 358 ProArgAsnAspSerLeuIleProAsnLeuGlyArg-----TyrIleThrGlu 373
Db 2110 CCGGGAAC-----AAGCGCGGTCAAGTCCAGCACGCGCCATCTCGCGGAT 2066
QY 374 GlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAsp 393
Db 2065 GTTCCACGTCCACATCCCGGATCACCAGTCGACAGCGTTGTATG-----CTCGGCGT 2012
QY 394 AspProTyrGlyLeu-----ProTrpTrpLysGluAlaValAla 406
Db 2011 CATCCGCGAGGATGTTCCTTAACACGACGTGTCACCGCGCTTGGCAGAGATAGAAAGC 1952
QY 407 Gln-----HisIleAlaLysAsnProThrAspAlaLeuProIleProPhe 421
Db 1951 CAGGTGTCGTGCTGCACACAGACGACCACTCTGCAACAGCCCATGAACACAG 1892
QY 422 ArgAspProGluProGlnValThrProPheThrGluGluHisProTrpHisThrGln 441
Db 1891 CAGAACAGGATACCTCTGTTTCGACCGAGCATTCGGCCCTTCTCCGCCCATGAACAC 1832
QY 442 IleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluVal-----AspSer 458
Db 1831 GTCCGG-----GCCCGGGGCCCTCTCTCTCTTCCGACAAAGG 1793
QY 459 ArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnLeuLeu 478
Db 1792 CGTGATACCTCTCGCGGTTTAAA-----AACATGCCACACACGCTAATAAAGGCC 1745
QY 479 ValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnPro----- 493
Db 1744 ATCTCCCTCTGGAGCTTGAAGTCTGCAGATTCAGCCCTACCGCGCGCGTAAATACCT 1685
QY 494 -----ThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMet 510
Db 1684 AACATCAACAAAGAGCAAGTCCGCCGCCCTCCAAAGCGGACACTCATCCGCGGCTC 1625
QY 511 MetAlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerPro 530
Db 1624 CTCGTC-----CACATCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 1586

QY 531 Gln-----PheMetAspProGlyLeuAlaLeuHisLeuAlaGly 543
Db 1585 TCTAAGCTCCTCCTCTCGACTCCACTGTGCTCGGCTCCTCGGC 1541
RESULT 11
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 5.05e+04 Length: 9025608
Score: 107.50 Matches: 153
Percent Similarity: 31.99% Conservative: 94
Best Local Similarity: 19.82% Mismatches: 268
Query Match: 3.27% Indels: 260
DB: 14 Gaps: 34
US-09-856-327-2 (1-618) x US-10-156-761-1 (1-9025608)
QY 15 SerMetGlnIleAsnGlyGlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAsp 34
Db 1793303 TCCGTGAAGAACAACCGGACAGCTGTCCCGCAGAGGCTGTCTCGCGGTTCGCG 1793244
QY 35 GlyValAspValPheIle---AlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeu 53
Db 1793243 GGCCTCGCGCCCTGACGGGCTGCGGACGCGGAGGGGAGGCGGACACGAGGACCTC 1793184
QY 54 CysValGlu---AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPhe 72
Db 1793183 TCCAAGAAGCGGCGCGC-----CCGATGAAGAAGTACTCGGGCGGGGACCATTC 1793133
QY 73 TyrAlaValAsnAlaGluGluGlyThrAlaValProTyr---ValProGlyTyrHisLys 91
Db 1793132 AAGGCCCAAGGCGCTCACCTTCAGTATCTGCACAAAGACATGCCACCTACCCGATG 1793073
QY 92 LysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAla 111
Db 1793072 AAGCGGACTCGCTGTCTTGAAGAGAGGTACCGGAGCGCAACCGCGGTC-----ACC 1793022
QY 112 LeuGlnGlnValSerValProValArgAsnGlnAsn----- 123
Db 1793021 CTCAGCGGACCTCGGTCCCGCCAGCGACTACGACAGAGAGCGAAGCCTGCTCATCGCG 1792962
QY 124 -----ValProThrLeuAspProGly-----Ala-TrpSerAl 134
Db 1792961 GCGGCGGACGCGCGCTTACTGATCCCCAAGACGTACCGCGGCGCAGGAGACCGGTCGTG 1792902

;; PRIOR FILING DATE: 1999-10-01
;; PRIOR APPLICATION NUMBER: US 60/139,650
;; PRIOR FILING DATE: 1999-06-17
;; PRIOR APPLICATION NUMBER: US 60/123,810
;; PRIOR FILING DATE: 1999-03-11
;; PRIOR APPLICATION NUMBER: US 60/102,748
;; PRIOR FILING DATE: 1998-10-02
;; NUMBER OF SEQ ID NOS: 72
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 24
;; LENGTH: 4725
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; NAME/KEY: CDS
; LOCATION: (3)...(4724)
US-09-940-316B-24

Alignment Scores:
Pred. No.: 0.692 Length: 4725
Score: 107.00 Matches: 134
Percent Similarity: 34.61% Conservative: 65
Best Local Similarity: 23.30% Mismatches: 217
Query Match: 3.26% Indels: 162
DB: 12 Gaps: 31

US-09-856-327-2 (1-618) x US-09-940-316B-24 (1-4725)

QY 23 ProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIle----- 40
| | | | | : : : : : | | | | | : : : : :
DB 1659 CCGCGCCGCGTCTCGCTCGCTGCTGCGGTGAGCGCACGACGCCCATCATCTT 1718
QY 41 ---AlaGly-----SerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
| | | | | : : : : : | | | | | : : : : :
DB 1719 GAGCAGCAGCGGTCAAAACGGGACCGGTGAGGACGGA-----GCGATCGAG 1766
QY 57 AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
| | | | | : : : : : | | | | | : : : : :
DB 1767 GCAGGA-----CCGTCGAGTAGGACCGGTGAGGCT----- 1799
QY 77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGlu 96
| | | | | : : : : : | | | | | : : : : :
DB 1800 GGACCGCTCCCGCGGCGCCGCTCAGCAGCGGC----- 1835
QY 97 PheGlnLysAspIleAspArgPheValAsnVal-Ile----- 108
| | | | | : : : : : | | | | | : : : : :
DB 1836 -----GAAGACCTTCCGCTCTGCTGCTGCGCGGTTCCCGGAGGCACTCGAGGACGAG 1889
QY 109 -LysGlyAlaLeuGlnGlnValSerValProValArg---AsnGlnAsnValProThrLe 127
| | | | | : : : : : | | | | | : : : : :
DB 1890 ATCGGGCCCTTCGCGGCTATCTCGACACCGCCCGCGGCTCGACCGGCGCGGTGGCG 1949
QY 127 uAsp-----ProGlyAlaTrpSerAlaProProGly-----SerSerAlaIleSe 142
| | | | | : : : : : | | | | | : : : : :
DB 1950 CAGACACTGGCCCGCGGTACGCACTTCACCCACCGGCGGTACTGCTCGGGGACACCGTC 2009
QY 142 rAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaValThrAr 162
| | | | | : : : : : | | | | | : : : : :
DB 2010 ATCGGGCGTCCCGCGGACGAGCGGACGAGCACTCGTCTTCTACTCCGGTCAGG-- 2067
QY 162 gGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHisProPome 182
| | | | | : : : : : | | | | | : : : : :
DB 2068 -----GCACCCAGCATCCCGCGATGG---GCGACCAACTCGCGCGCGCTTCCCGGTG 2117
QY 182 tGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAspAspLy 202
| | | | | : : : : : | | | | | : : : : :
DB 2118 TTCGCGGATGCTGGCAGCAGCGCTCCGAGCGGTTCGACGACCGCCCGACCGACCGCCC 2177
QY 202 sGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLysGluPh 222
| | | | | : : : : : | | | | | : : : : :
DB 202 sGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLysGluPh 222
| | | | | : : : : : | | | | | : : : : :
DB 3018 GGGT-----CGCGCGCGCGGTGTTTCACGGGTC 3045

DB 2178 ACAC-----GGAGCCAGCACACGCTC-TT 2200
QY 222 eAspGluSerIleArgHisThrLeuValLeuArgSerLeu----- 235
| | | | | : : : : : | | | | | : : : : :
DB 2201 CGCCACACAGCGGCTTACCGCCCTCTGAGGTCCTGGGACATCAGCGCGACCGCT 2260
QY 236 -----GlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCy 253
| | | | | : : : : : | | | | | : : : : :
DB 2261 CATCGGCACATC-GCTCGGCGAGATCACCAGCGGTACGCGCGGGATCTGTGCTGCTG 2319
QY 253 sHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAla-GluAsnLeuPheH 273
| | | | | : : : : : | | | | | : : : : :
DB 2320 ACAGCGCTGCAC-----CCTGAT-----CACCACGCGTCCCGCTCATGC 2361
QY 273 isSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgC 293
| | | | | : : : : : | | | | | : : : : :
DB 2362 ACAGCGTT-----CCGCGCGCGCGCGGTGAGTCCGCTGACACGAGGAGG 2412
QY 293 ysThrArgLeuAlaLeuThrGlyGlyTyrGlu-LysLysIleGlyAlaAlaGluValArg 312
| | | | | : : : : : | | | | | : : : : :
DB 2413 AGGCGCGTCAAGCGCTGCGCGCGGTGAGATCGCGCGGTCTTTCGGCC- 2463
QY 313 AsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLysVal 332
| | | | | : : : : : | | | | | : : : : :
DB 2464 -----CGCACTCCGTCG----- 2475
QY 333 TyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSer 352
| | | | | : : : : : | | | | | : : : : :
DB 2476 -----TCCTCTCGGCGCAGGAGGCGCGGTGCTGACGTCGACGCGGTCTCGGCA 2526
QY 353 GlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrIleThr 372
| | | | | : : : : : | | | | | : : : : :
DB 2527 TCCACACCGCTCCCGCGCGCGCGCGGTGAGTCCGCGCACATGAAACCC-CTGCGCC 2585
QY 373 GluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArg 392
| | | | | : : : : : | | | | | : : : : :
DB 2586 GCCAGGTGCTCGCGCACCACTCGCGAGCTCTTACGACCGCGCCACACCGCATCCG 2645
QY 393 AspAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGln----- 407
| | | | | : : : : : | | | | | : : : : :
DB 2646 AACACCCACCCACCGCGAGTACTGGCGCGAGGAGTCCGCAACCCCGTCTGTTCAC 2705
QY 408 ---HisIleAlaLysAsnProThrAspAlaLeuProIle-----ProPheArgAsp 423
| | | | | : : : : : | | | | | : : : : :
DB 2706 GCCACACCCACCGGTACCC- -GACCGCTGTCTGTCGAGATCGGCGCCGCGCAGGAC 2762
QY 424 ProGluProGlnVal-----ThrThrProPheThrGluGluHisPro 437
| | | | | : : : : : | | | | | : : : : :
DB 2763 CTCTCACCGTGGTTCGAGCGCATCCCTGCGAACGCGACGCGGAGGTGCACGCG 2822
QY 438 TrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValAlaGlyProGluValAsp 457
| | | | | : : : : : | | | | | : : : : :
DB 2823 CTGCACACCGCGCTCGCGCG- -CTCTTACACGCGCGCCACG-----CTCGAC 2870
QY 458 ---SerArgValIle-ValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAs 476
| | | | | : : : : : | | | | | : : : : :
DB 2871 TGGTCCCGCATCTCTCGCGGTGCTTCGCGGACGACCCCTGACGTCCTCTGACGCGTC 2930
QY 476 nLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSer---MetProGlnProThrPh 495
| | | | | : : : : : | | | | | : : : : :
DB 2931 C-----AGCGCGTCCCTACTTGGATCGATCGGTCTCCCGCGGCA----- 2970
QY 495 eArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCy 515
| | | | | : : : : : | | | | | : : : : :
DB 2971 -----CGCGCGACTCGGGCCACC----- 2988
QY 515 sGluValAlaSerAsnLeuGlyTyrLeuProThrSerProGlnPheMetAspPr 535
| | | | | : : : : : | | | | | : : : : :
DB 2989 -----CCGTCTCGCGCACCGGAGTCGCGCTCGCC 3017
QY 535 oGlyLeuAlaLeuHisLeuAlaGlyThrArgIle 547
| | | | | : : : : : | | | | | : : : : :
DB 3018 GGGT-----CGCGCGCGCGGTGTTTCACGGGTC 3045


```

QY 223 AspGluSerIleArgHisThr-----LeuValLeuArgSerLeuGlnAspAlaTyr 239
Db 730 GACCCGATGGCGTGCACACCGGTACTCGATCACCGT-CGCGCCCTCGATGACGCT-- 785
QY 240 LysAspArgGlnArgIlePheArgProLeuAlaCysHisArgLeuLysAsnAla 259
Db 786 ---GACCGACCGCGAGTA-----CCAGCGCT---GCGCGACATCGGTATCGCATCAT 833
QY 260 ProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAsp 279
Db 834 CGGTGAGTTCGTGTGCACAC-----CGCGCGTGCACATCCAGTTCGCGTGAACCC 887
QY 280 LysGlnLys-LysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAla----- 297
Db 888 CGACGACGCGCGATCATGTCATGAGATGAACCGCGCGTGTCCCGTCTCGTCCGCGCT 947
QY 298 -----LeuThrGlyGlyTyrGluLysLysIleGlyAlaAla----- 309
Db 948 CGCCTCCAAGCGACCGGCTTCGCGATCGCGAAGATCGCGCGAAGCTCGCGTGGCTA 1007
QY 310 -----GluValArgAsnLeuLeuAlaThrArgAsnProSerSer-----G1 323
Db 1008 CACGCTCGACGAGATCCCAACGACATCACCGAGAGACCCCGCGCTCTTCGAGCCGAC 1067
QY 323 nLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnPr 343
Db 1068 GCTCGAC---TACGTGGTGTGAAGGCC----- 1094
QY 343 oGlnIleLeuTyrAsn-----SerGlyPheSerGlyLeuGlnValThrProArgAs 360
Db 1095 GCGCTTCGCTTCGAGAAGTTCCCGAGCGCCGACTCGACCTTCACACACCATGAAGTC 1154
QY 360 nAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysgl 380
Db 1155 GGTCCGCGAGGCCATGGCGCATCGCGGGAACCTCACCGAGGCG-----CTGCA 1202
QY 380 nIleValLeuArgGln-----GluPheValAspSerValAr 392
Db 1203 GAAGCGCTGCGGTCCTCGAGAAGAAGGCTCGCAGATTCAGCTTCGTGCGGAGCGGG 1262
QY 392 qAspAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGlnHisIleAlaLysAs 412
Db 1263 CGAC-----AAGGCCCTTCCTCTGGAGGAGCGGCTCGG 1295
QY 412 nProThrAsp-----AlaLeuProIleProPheArgAspProGluProgl 427
Db 1296 TCCACCGACGAGCGCATCAACTCGGTATCGACGACCCATCCGCGCGGCGGACGCCGA 1355
QY 427 nValThrThrProPheThrGluGluHisProTyrThrHisThrGln-----IleHisAr 444
Db 1356 GGAGTCTTCGAGGCGACGACAGATGACCCCTGGTTCGTTCGACACAGCTTCTCTGATCAA 1415
QY 444 qAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgValIleValAsple 464
Db 1416 GGAATCGGGGACGAGCTGGCGCGCGACAAAGCTGGACCGCGAGCTGCTCGCGAGGCG 1475
QY 464 uArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspVa 484
Db 1476 CAAGCGGACGCGCTTCTCCGAGGTCCAGATCCCGAGATCCCGGCGTCCGCGAGGACGT 1535
QY 484 l 484
Db 1536 c 1536

```

RESULT 15

US-10-014-717-1/c

; Sequence 1, Application US/10014717

; Publication NO. US20020192778A1

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

```

; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1

Alignment Scores:
Pred. No.: 49.6 Length: 68750
Score: 106.00 Matches: 106
Percent Similarity: 33.26% Conservative: 47
Best Local Similarity: 23.04% Mismatches: 184
Query Match: 3.23% Indels: 123
DB: 13 Gaps: 25

US-09-856-327-2 (1-618) x US-10-014-717-1 (1-68750)
QY 116 SerValProValArgAsnGlnAsnValProThrLeuAspProGly-----AlaTrp 132
Db 29586 TCGGTCTCTCGGACATAGCTCGGTGACGACGACTCTCGCCCTCGATCGGGTGGTGG 29527
QY 133 -----SerAlaPro-----ProGlySerSerAla 140
Db 29526 GCGGGGCTCGTTCGGGATCCAGGTCCACGAGCCGCCGCCCGCGGATGCTCGAGC 29467
QY 141 IleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
Db 29466 GCGGCCACCGGCCCTCGCCCAT-----AACGCGCTCGACAAGGGCGG 29422
QY 161 ThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHis-Pr 180
Db 29421 ATCGCAGGC-----TCGTCCGCAACGATGATGCC 29392
QY 180 oProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAs 200
Db 29391 CCCC---GGGTACGACCCAGAGTCGGGCGCAACAAGACACGCGTCTCAGAAACCGAGCC 29335
QY 200 pAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrly 220
Db 29334 AAGCCGA-----GCACCGCGCGG 29317
QY 220 sGluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrIly 240
Db 29316 GTAGCAGCGGGTTCGCTCGC-----CGATCTCATGATCGACGCCCTCC 29272
QY 240 sAspArgGlnArgIlePheArgProLeuPro-----LeuAlaCysHisArgIle 256
Db 29271 GCACCGCAGCAGCG-----CGTCCAGACCCACGAGTAGACACTACCTGCCAATCGCTT 29218
QY 256 uLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyr 276
Db 29217 C----- 29217
QY 276 rAsnAspAspLysGlnLysLysLeuPheThrLeuThrAsnHisArgCysThrArgIle 296
Db 29216 ---GACCGCGCGGACGCTCGG-----TCACCAGTTCGCGGTGCGGAT 29176
QY 296 uAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAl 316
Db 29175 GTCTCTCGCGCGCATGGAGCAGCAGCATG-----GAAGTCCACGCTCTCCACAGC 29125
QY 316 aThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAl 336
Db 29124 GCTGACGCGACCGCTCTCGCGACTCCACCTTAT-----CCGCCAATACC 29080

```

```
QY      336 aSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerglyPheSerglyLeuGlnVa 356
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
29079   AGCAGCTCCCGCGCTCGCTCTCTGATT-----TCTGGA-----GGCTCGGAGT 29032

QY      356 lThrProArgAsnAspSerLeuIleProAsnLeuGly-----ArgTyrIleThrGl 373
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
29031   ATCTCCGGCCAGTCCACCGATAGAACACGAGCTTGGTCCGATCCGCGCTCGTGGCGA 28972

QY      373 uGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAs 393
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28971   CGGCTGTTCAGGCTCGATTCGGGCCAGTACCCCTGCCCTGCCACGGATAGGTCCGCGAGC 28912

QY      393 pAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGlnHisIleAlaLysAsnPr 413
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28911   GGCACCCCGCGCCAGCGTGGGAAGACGCCCGCCAGCTGACCGAGCGCGCGCGCGCC 28852

QY      413 oThrAspAlaLeuProIleProPheArgAspProGluPro-----GlnValTh 429
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28851   CACA---GCCTCGCCAGCGCTTCGAGCACCCCGCAGCTCTCGCGCCCGCGCGCAAC 28795

QY      429 rThrProPheThrGluGluHisProTyrHisThrGlnIleHisArgAspAlaPheSerTy 449
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28794   GAGCGCAGCAGCGTCCGCTCCCTCCGCGCAGGAGGCTG---GCAACAGCCCGAGAGC 28738

QY      449 rGlyAlaValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAl 469
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28737   GTCGGCTTCGGGCCCA-----CTTCGACGAACGTGCCCGCACCGGCTTCGTGCAGC 28687

QY      469 aThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSe 489
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28686   GCCTTCACCCCGCTCCGCGA-----AG 28666

QY      489 rMetProGlnProThrPheArgTyrArgProSerThr---AlaSerAsnValArgAlaAr 508
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28665   CGCACCGCTCCGCGA---CGTGGCGCACCCAGTACCCCGCGCGCTCAGCTCTCCCGG 28609

QY      508 gLysMetMetAlaAspMetCysGluValAlaSerAsnLeu---GlyGlyTyrLeuProTh 527
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28608   ACCACCTTCCCGCTCAGGTGCTCACCGAGCAAGCTTGGCGCGCGGTACGTACCGGAC 28549

QY      527 rSerProProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArg 546
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28548   GCCGCCACCCCGCGCACTCTCCAGCATCGGTTCCTCAGCGCGGAGTGGGACGCGT 28491
```

Search completed: September 24, 2003, 15:19:29

Job time : 23190 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 05:57:01; Search time 104 seconds
(without alignments)
2622.834 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSPEQLRDPYRSMQING.....IINTLKGTGDKNTGHRNL 618

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL-frame+ p2n model -DEV-xlh

-Q/cgn2_1/USPTO_spool/US09856327/runat_23092003_153420_15557/app_query.fasta_1.775
-DB-issued_Patents_NA -QWtr=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -SPART=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMF=ptto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856327@cgn1.1.56.0runat_23092003_153420_15557 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1232.5	37.5	1869	1 US-08-734-925-1	Sequence 1, Appli
2	1210	36.8	1869	3 US-09-305-381-1	Sequence 1, Appli
3	1091	33.2	1701	3 US-09-023-731-2	Sequence 2, Appli
4	198	6.0	4665	4 US-09-297-937C-12	Sequence 12, Appli
5	191.5	5.8	1845	4 US-09-297-937C-8	Sequence 8, Appli
c 6	148	4.5	4403765	3 US-09-103-840A-2	Sequence 2, Appli
c 7	148	4.5	4411529	3 US-09-103-840A-1	Sequence 1, Appli
c 8	146	4.4	1944	4 US-09-252-991A-9081	Sequence 9081, Ap
c 9	128.5	3.9	42225	4 US-08-311-731A-131	Sequence 131, App
10	126	3.8	1611	4 US-09-252-991A-11302	Sequence 11302, A
11	124	3.8	2576	3 US-09-265-108-1	Sequence 1, Appli
12	124	3.8	2576	3 US-09-479-264-1	Sequence 1, Appli

13	122.5	3.7	2106	4 US-09-252-991A-11428	Sequence 11428, A
14	122.5	3.7	2400	3 US-08-930-001-1	Sequence 1, Appli
15	122.5	3.7	3009	4 US-09-252-991A-11409	Sequence 11409, A
16	115.5	3.5	20235	1 US-07-642-734C-3	Sequence 3, Appli
17	115.5	3.5	20235	3 US-08-439-009A-3	Sequence 3, Appli
18	115	3.5	1758	4 US-09-252-991A-1644	Sequence 1644, Ap
19	114	3.5	4403765	3 US-09-103-840A-2	Sequence 2, Appli
20	114	3.5	4411529	3 US-09-103-840A-1	Sequence 1, Appli
c 21	113.5	3.5	1932	4 US-09-252-991A-7230	Sequence 7230, Ap
22	113.5	3.5	2898	4 US-09-252-991A-7467	Sequence 7467, Ap
23	109.5	3.3	1863	3 US-09-199-229-1	Sequence 1, Appli
24	109.5	3.3	1863	3 US-09-443-087-1	Sequence 1, Appli
25	109.5	3.3	1863	4 US-09-687-298-1	Sequence 1, Appli
26	108.5	3.3	1848	3 US-08-333-802-1	Sequence 8, Appli
27	108.5	3.3	2346	3 US-09-193-503B-8	Sequence 8, Appli
28	108.5	3.3	2346	4 US-09-415-838-8	Sequence 8, Appli
29	108.5	3.3	2788	4 US-09-347-878-45	Sequence 45, Appli
30	108.5	3.3	3026	3 US-08-981-729-6	Sequence 6, Appli
31	107.5	3.3	3396	3 US-08-974-549A-642	Sequence 642, App
32	107	3.3	4725	4 US-09-410-551B-24	Sequence 24, Appli
c 33	106.5	3.2	1230025	4 US-09-198-452A-1	Sequence 1, Appli
34	106	3.2	1773	4 US-09-134-001C-1678	Sequence 1678, Ap
c 35	106	3.2	68750	3 US-09-335-409-1	Sequence 1, Appli
c 36	106	3.2	68750	4 US-09-568-102-1	Sequence 1, Appli
c 37	106	3.2	68750	4 US-09-567-969-1	Sequence 1, Appli
c 38	106	3.2	68750	4 US-09-568-480-1	Sequence 1, Appli
c 39	106	3.2	68750	4 US-09-568-486-1	Sequence 1, Appli
c 40	106	3.2	68750	4 US-09-568-472-1	Sequence 1, Appli
c 41	106	3.2	68750	4 US-09-567-899-1	Sequence 1, Appli
c 42	105	3.2	1080	4 US-09-252-991A-11464	Sequence 11464, A
43	104.5	3.2	32798	4 US-09-604-694B-1	Sequence 1, Appli
44	104.5	3.2	34303	2 US-08-735-609-4	Sequence 4, Appli
45	104.5	3.2	34303	2 US-08-735-609-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-734-925-1
; Sequence 1, Application US/08734925
; Patent No. 5712139
; GENERAL INFORMATION:
; APPLICANT: NISHIMURA, IKURO
; APPLICANT: OKADA, KIMI HARU
; APPLICANT: MINAMIHARA, TOMOYUKI
; APPLICANT: KAWAI, GENSIRO
; APPLICANT: KOYAMA, YASUJI
; APPLICANT: SUZUKI, MASARU
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBOLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/734,925
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/568,428
; FILING DATE: 06-DEC-1995
; ATTORNEY/AGENT INFORMATION:

1594	Db	CTTGGGTCACCCACAGTTCATGGAGCCCGGCTTGTCTGCACTTGTGGGACGCAC	1653
546	Qy	ArgIleGlyPheAspLysAlaThr-----ThrValAlaAspAsnAsnSerLeuValTrp	563
1654	Db	CGCATGGCTTCGACGAGAGCGGCAAGTGTGCGTCGACACCGACTCACGGCTTTC	1713
564	Qy	AspPheAlaAsnLeuTrpValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsn	583
1714	Db	GGCTTCAGGAACCTCTTCCTCGCGGCTCGGGGAACATCCCAACCGGTACCGCGCAAC	1773
584	Qy	ProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsnThrLeu	603
1774	Db	CCAGCGCTACCGCAATGTCGCTTGCATCAAGACTCAAGCTGCAGATACATC-----AAG	1824
604	Qy	LysGlyGlyThrAspGlyLysAsnThrGlyGluHisArgAsn	617
1825	Db	AAGAAGCTTCGAGCGGAGCCCGCAACCCCGTGAAGCACCACCAAC	1866


```
QY 171 TrpThrCysSerThrProArgIleHisProProMetGluSerLeuProGly----- 187
Db 379 TGGTCAGCGGTACATTTCCGGTGCACCGGTGAAGCTGCGCAGCAATATGAA 438
QY 188 -----IleGlyArgProLysLeuSerAsnAspProAlaGluAsp-----LysGlu 203
Db 439 GCGGTTACGGCAAAACTTTATCCCGAAGGATGACGATTACAGGATTTCGGGCGTCAGC 498
QY 204 TrpAsnGluLeu-----TyrSerGluAlaGluArgLeuIleGlyThrSer----- 218
Db 499 TATAACCAAGTGAACCTTCTTCGATCAGCGCGAGAAAGTCTTTGGTACTCTCGGCGAGT 558
QY 218 ----- 218
Db 559 GCCTGGACCATCAAGCAAGATGATCGCAAGGAGAAAGCGCGCAACTTTTACGCGCCG 618
QY 219 -----ThrLysGluPheAspGluSerIle-ArgHisThrLeuValLeuArgSerLeuG1 236
Db 619 GACCGCTCCAGCGACTTCCCGTGCAGCGCACAGAGCGGACTTACTCGGCGCAGCTGTTT 678
QY 236 nAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLe 256
Db 679 GCCAGCGCGCAGAGTCGGTGGGTATCATCCGTACGATA-----TGCCATCGGCC 729
QY 256 uLysAsnAlaProGluTyrValGluTrpHisSerAlaGlu----- 269
Db 730 -AACACCTCAGGTCGCTACCAACACACTACGGCGCACAGATGGCGCCGTGCAACTCTG 788
QY 270 -----AsnLe 271
Db 789 CGGTATTGCGAGCGGTACCGCTGCTACATGTATTCAAAGCGTCGCCTAACGTGAACAT 848
QY 271 uPheHisSerIleTyrAsnAspLysGlnLysLeuPheThrLeuLeuThrAsnHi 291
Db 849 CCGTCCGCGCTG-----CGTCAGAGCGCAAGTTTCAGCTGCGTAAACACGC 896
QY 291 sArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysIleGlyAla----- 308
Db 897 ATATGTGTTCGGGTCAATCTGACCGCGCACAAACGCGCCACTGGCTGACCTATCT 956
QY 309 ----AlaGluValArgAsnLeuAlaThrArgAsnProSerSerGln-----LeuAs 325
Db 957 CGATGTCAGGTCGGAAGTGTG-----CAGCCTGCGGATCTGCTGATCTGTCGTGC 1007
QY 325 pSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnPro---G1 344
Db 1008 AGCGTTCACGTTCCACAACTGCACCTGATGCTGCTCCGGTATCGCCAGCGATATA 1067
QY 344 nIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeu11 364
Db 1068 CCGGATCCTAACGAAGGTGTGGTCCGCGCTAACTTCGCTTATCAGAACAATCTCGACGCT 1127
QY 364 eProAsnLeu--GlyArgTyrIleThrGluGlnProMet-----AlaPheCysG1 380
Db 1128 GAAAGCGCTTTCACAAACAAACACCCTACTAACCGCTTTATCGGTGCGGCGGCGCAGG 1187
QY 380 nIleValLeuArgGlnGluPheValAspSerValArgAspProTyrGlyLeu----- 398
Db 1188 GTGGCGGTGGTGAATTTCACACCCGACAACTTCGACCGCGCCGCTACGGCTTCGTGCG 1247
QY 399 -----ProTyrTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAl 416
Db 1248 TGGCTCGCCATTCTGG-----GTGAACCGAGCGGGTACCAAAACCGGTTTCGGG 1295
QY 416 aLeuProIleProPheArgAspProGluProGlnValThrThrProPheThrGluGluHi 436
Db 1296 TCTGCCGACCCCAAGGCGACCGCGAAC----- 1323
QY 436 sProTyrHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluVa 456
Db 1324 ----TGGGCGAGCGCATGGNAA-----GCCGCGGTGCGGATACCTA 1361
QY 456 lAspSerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAs 476
```

```
Db 1362 CAACCACCATATTTTCGATGGATGCCAC-----GGTGGCACCAAGTCATACCGCGCTAA 1415
QY 476 nLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetPro-----GlnProTh 494
Db 1416 CTACCTCGATCTCGATCCGAACTACAAAAATGTCTACGGCAGCGCTGCTGCGTATGAC 1475
QY 494 rPheArgTyrArgProSerThrAlaSerAsnValArg---AlaArgLysMetMetAlaAs 513
Db 1476 CTTTGACTGGCAGGAC-----AACGACATCAGGATGCGCAGTTTATGTCGCGAA 1526
QY 513 pMetCysGluVal-----AlaSerAsnLe 521
Db 1527 GATCGCGAAATCACCGAGGCCATGAATCCGAAGATGATCATCGCGCGCTAAGGGACC 1586
QY 521 uGlyGlyTyrLeuProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHisLe 541
Db 1587 GGTACCACCTTCATACCAACCGGTATCAA-----ACCACGCATAT 1628
QY 541 uAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnSerLe 561
Db 1629 GAGCGCGGGCGCATCATGGTGAAGATCCGAAACACGAGCAGTGAACCGTTATTGCA 1688
QY 561 uValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheG1 581
Db 1689 GAGCTGGGATGTCGCAACGTGTTTGTGCGGGTGCCTCCGCGTTCGCCGAGGCTCTGG 1748
QY 581 yGluAsnProThr 585
Db 1749 CTACAACCCGACC 1761

RESULT 6
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 2.09 Length: 4403765
Score: 148.00 Matches: 144
Percent Similarity: 32.37% Conservative: 80
Best Local Similarity: 20.81% Mismatches: 206
Query Match: 4.51% Indels: 264
DB: 3 Gaps: 31

US-09-856-327-2 (1-618) x US-09-103-840A-2 (1-4403765)
QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
Db 3819533 GAGCTCTGATTATTGATCGGTTTCGGGCGGAGCGTCAACCGCTGCGGCTGACGCGAA 3819474
QY 57 AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
```

Db 3819473 AAGGCTACCGGGTGGCGTCTCGAGCGCGC----- 3819441
Qy 77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGlu 96
Db 3819440 -----CCTCGGTTCTCCGACGAGGAG 3819420
Qy 97 PheGlnLys-----AspIleAspArgPheValAsnValIleLysGlyAlaLeuGln 113
Db 3819419 TTCGCCAAGACGCTCGTGGGACCTCGCAAGTTTCTC----- 3819384
Qy 114 GlnValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSer 133
Db 3819383 -----TGGGCG 3819378
Qy 134 AlaProProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGlu 153
Db 3819377 CCGAGTTGGCTGCTATGGCATT-----CAACGTATCCACCGCTCGCAACGTG 3819327
Qy 154 AsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrCys 173
Db 3819326 ATGATCCTGCGCGCGCC-----GGGTGGCGGGGCTCCCTCAACTAC---GCC 3819279
Qy 174 SerThrProArgIleHisProProMetGluSerLeuProGlyIleGlyArgProLysLeu 193
Db 3819278 AACACGTGTACGTGCGCGCG----- 3819258
Qy 194 SerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeu----- 207
Db 3819257 -----GAGCCGTCTTCGCGCACGACGAGTGGTCACATCACCGACTGCGTGGCGAG 3819204
Qy 208 -----TyrSerGluAlaGluLeuIleGlyThrSerThrLys----- 220
Db 3819203 CTGATCGCGCACTACCAAGCAGCGCAGCGGATGTCGGGT-GGTGCAAAACCCGACCTT 3819145
Qy 221 -----GluPheAspGluSerIleArgHis--- 228
Db 3819144 CACCGACGCGCACCGCATCTGCAAGAGGTGCGCGACGAGATGGGTTTGGCGACACCTG 3819085
Qy 229 -----ThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAsp 241
Db 3819084 GGTGCCGACCCGGTTGAGTGTCTTCGGCCCCGCGACGCGCACCAAGCGCCGCGCAAGAC 3819025
Qy 242 ArgGlnArg-----IlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAla--- 259
Db 3819024 CGTGCCCGATCCCTACTTTCGGGTGCGCGCGCGCGCACCGCTGCTGGAATGGG 3818965
Qy 259 ----- 259
Db 3818964 CTGCTGTATGACGGCTGTGCGCCAGCGGTGCCAAGACACCTGTGTGAAAACTACCTTGG 3818905
Qy 260 ProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspAsp 279
Db 3818904 CCTCGCGAATCAGCTGGCGCGCAAGTAT-----TCCGATGAC 3818866
Qy 280 -LysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuTh 299
Db 3818865 CACGGTGAAG----- 3818856
Qy 299 rGlyGlyTyrGluLysLysIle---GlyAlaAlaGluValArgAsnLeuLeuAlaThrAr 318
Db 3818855 -----GGGTTTGGCGGCGCTCGACGCGGTGTGGAGGTGCGCACAGTG----- 3818811
Qy 318 gAsnProSerSerGlnLeuAspSerTyrIle-----MetAlaLys 331
Db 3818810 -----CGAACCGCAGTGTGGCTGCGCGGGATCGGCGCACCTTCACCGCAAC 3818764
Qy 331 sValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyr----- 347
Db 3818763 GCAGCTGTGTGGCGCGCGGCGCACCTGGGGGACCCAGCATCTGCTGTTCAAGATGCGGGA 3818704
Qy 348 -----AsnSer-- 349
Db 3818703 TAGGGGTGCGTCCCTGCTCTCTCGAAGCGATGGGTGCTCTTACCGCGACCAACTCAGA 3818644

Qy 350 -----GlyPheSerGlyLeuGlnValThrProArgAsnAsp----- 361
Db 3818643 ATCGATCGTCGGGCGCCGACCTTGAAGGTCAATCCGCGACCTGACCCACCGCGGT 3818584
Qy 362 -----SerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAl 377
Db 3818583 GCGCATCACGTCGTCGATTCCACCGACGCGCGCACCCACATC-----GAACCGGTGCG 3818530
Qy 377 aPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAspAsp----- 394
Db 3818529 CTACGGCAAGGCTCCACGCGATGGGGTCTGTCACACCTTGTATGCCGACGCGCTCCGG 3818470
Qy 395 -ProTyrGly-----LeuProTyrTrpLysGluAlaValAlaGlnHisIleAlaLysAs 412
Db 3818469 TCCGACGGGCACCGCATGTCGCGCTGGAGGCGAG---TTGCTGCAGACGCGCAGTCAGGA 3818413
Qy 412 nProThrAspAlaLeuProIleProPheArgAspProGluProGlnValThrThrProPh 432
Db 3818412 TCCG-----CGCGGCAC-CATCCGATGCTCAATCCCGCGC 3818378
Qy 432 eThrGluGluHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaVa 452
Db 3818377 AATGAG-----CGAGCCACCGTATCGCGCTGTCATGACGACCC 3818336
Qy 452 lGlyPro-----GluVal-AspSerArgValIleValAspLeuArgTrpPheGlyA 469
Db 3818335 TGGACAACCTCATCCACCGTTCCAAAGCGCGGGAAGTGGGTATCGCTGCTACTCGA 3818276
Qy 469 laThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAspGlyTyrS 489
Db 3818275 GC-----AAGCAGGACACG 3818261
Qy 489 erMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaAArgL 509
Db 3818260 GCGAGCCGAACCTCATATGG-----ATCCCGATCGGCAACCGAGTCACCGCTC 3818213
Qy 509 ysmetMetalAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerP 529
Db 3818212 GCATCGCGCGCAAAATCAGCGCGTGGCGCGGCGCACCTGGGTGCTGTTCAACATCC 3818153
Qy 529 roProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyP 549
Db 3818152 CG-----CTCACCGCGCATTTCTTCGCGCGCGCGGTGATGTCGCG 3818114
Qy 549 heAspLysAlaThrThrValAlaAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeu 569
Db 3818113 ACGACCCGCGACGCGGTATCATCCCTATCACCGGCTCTATGGCTATCCGACGCTGT 3818054
Qy 569 yrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerM 589
Db 3818053 ACGTGGTTGACGGTGGCGGATCTCGGCGAACCTGGGTGTCAACCGCTGCTGTCCATCG 3817994
Qy 589 etCysHisAlaIleLysSerAla-ArgSerIleIleAsnThrLeuLysGlyGlyThrAsp 608
Db 3817993 CCGCTCAAGCGCGCGCGCGCTGCTGT-----GGCCGAACAAGG 3817952
Qy 609 GlyLysAsnThrGlyGluHisArgAsnLeu 618
Db 3817951 CGAAACCGACCGCGCGCGCGCGGTG 3817922

RESULT 7

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00


```
Db 3827288 ACGACCCCGACGCGCTCATCCACCCCTATCAACCGGCTGTATGGCTATCCGACGCTGT 3827229
Qy 569 yrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProHisGlnArgGluPheGluA 154
Db 3827228 ACGTGGTTCAGCGTGGCGGATCTCGGGGAACCTGGGTGTCAACCCGTCGCTGCCATCG 3827169
Qy 589 etCysHisAlaIleLysSerAla-ArgSerIleIleAsnThrLeuLysGlyGlyThrAsp 608
Db 3827168 CCGCTCAAGCCGAGCGGCGCTCGCTGT-----GGCGGAACAAGG 3827127
Qy 609 GlyTysAsnThrGlyGluHisArgAsnLeu 618
Db 3827126 GCGAAACCGACCGCGCGCCACCGCAGGCTG 3827097

RESULT 8
US-09-252-991A-9081/c
; Sequence 9081, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9081
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (4)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-9081

Alignment Scores:
Pred. No.: 8.91e-06 Length: 1944
Score: 146.00 Matches: 147
Percent Similarity: 30.98% Conservatives: 68
Best Local Similarity: 21.18% Mismatches: 271
Query Match: 4.45% Indels: 210
DB: 4 Gaps: 24

US-09-856-327-2 (1-618) x US-09-252-991A-9081 (1-1944)
Qy 15 SerMetGlnIleAsnGlyGlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAsp 34
Db 1913 TCCCTTCGCGCGGTGGCGATTTCGGCGGAGGGCTGAACATGGCGGACGACACTCAA 1854
Qy 35 GlyValAspAlaPheIleAlaGly-SerGlyProIleGlyAlaThrTyrAlaLysLeuCy 54
Db 1853 CCGGTGGACGGGTGGTGTCTCGGCTTTCGGTGGACCGGAGGATCTCTGGCCAGGAGCT 1794
Qy 54 sValGluAla-GlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrA 74
Db 1793 GACCGAGCGCGGGCTCAACGTGTGGCTGGAGCGGCGGAGAACCGCATACCTATC 1734
Qy 74 LaValAsnAlaGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnG 94
Db 1733 CCGAGCGCGCTACCCCAACACCTC-----GAGC 1704
Qy 94 luIleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnG 114
Db 1703 AGCTGACCTACAACACCGCGGCAAGCTGTTCAGAAATCTCTCCAAGACCGCTG---- 1648
Qy 114 lnValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerA 134
```

```
Db 1648 -----
Qy 134 laProProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluA 154
Db 1647 -----AGCATCCGACGGATCAACGACACCGCGCTGCCCTACCGCC 1605
Qy 154 snLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisThrPheCys 174
Db 1604 AGCTCTCGCGCTTCTTCCCGCGACGCGTGGTGGCGCGGCTGCACTGTGTCGCGC 1545
Qy 174 erThrProArgIleHisProMetGluSerLeu-ProGlyIleGlyArg----- 190
Db 1544 TGCATTTCCGGATCATGCCGAGGAACCTGCCGACGCAATACGAGAACGCTACG 1485
Qy 191 -----ProLysLeu-----SerAsnAspProAlaGluAspLysGlu 203
Db 1484 CCAAGAAATTCATCCCGAAGCATGACCATCCAGGACTACGGGTGCTACGAGAGC 1425
Qy 204 TrpAsnGluLeuTyrSerGluAlaGluArgLeuGlyThrSerThrLysGluPheAsp 223
Db 1424 TGGAGCGCAC-TTCGATTTCGCGGAGAAGTCTTCGCGCACTCC----- 1381
Qy 224 GluSerIleArgHisThrLeu----- 230
Db 1380 ---GGCACCGCGCATACGGTCAAGGGCCAGGTGTGGCAAGGGCAATCCGTTCCGCC 1324
Qy 231 -----ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArg 244
Db 1323 GACCGCTCGGACGACTTCCCGCTGCCGCACTGCCGAGGTGTACTCC-----CGCAA 1270
Qy 245 IlePheArg-----ProLeuProLeuAla--- 252
Db 1269 CTGTTCCGCAAGCGCGGAGGAACCTCGGCTGCAACCTTACGACCTCGCGCGCCAAC 1210
Qy 253 -----CysHis 254
Db 1209 GCTTCGGGCGCTGGACAAACCCCTACGGGTGCAGATGGGGCCGTGCAACTTCGCGCG 1150
Qy 255 ArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSer 274
Db 1149 TTCGACGCGCTACGCTGTACATGTATTCGAAGGCTCGCCGAACCTG-----AAC 1096
Qy 275 IleTyrAsnAspAspLysLysLeuPheThrLeuLeuThrAsnHisArgCysThr 294
Db 1095 ATCCTCCGCGCGCTGCGGACAGCCGCTGTTCGAACCTGCGCGCACTGCAACGTGCTC 1036
Qy 295 ArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAla-----AlaGlu 310
Db 1035 AAGGTCAACCTCGATAGCGATGGCAGCGGCGGCGGCGGCTGACCTACGTCGACCGCCAG 976
Qy 311 ValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAla 330
Db 975 GGGCGGAGATCGTGACCG-----GGG 952
Qy 331 LysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGly 350
Db 951 AAGCTGGTATCATCAGTGCCTCCAGTTCACAACGTCGCGCTGCTGCTCTCCGCGC 892
Qy 351 PheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyr 370
Db 891 ATC---GGCAACGCGCTACGACCCCGCGGCGGCGGCTG-----GTGGGGAAC 841
Qy 371 IleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSer 390
Db 840 TTCGCTTACCAAGACATGGCC-----ACCATCAAGCGCTTTTCGACAAGGAC 793
Qy 391 ValArgAspProTyr----- 396
Db 792 GTGATACCAACCCCTTTCGCGCACCGCGGCGGCGGCTGGCGGTGGAGACTTCAAC 733
Qy 397 -----GlyLeuProTyrTrpLysGlu 403
Db 732 GCCGACAACTTCGACCCAGCGCTGGGCTTCGTGCGCGGCTCGCCGATGTGG----- 679
```

```
QY 404 AlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAsp 423
Db 678 -----GTCAACAGCGCGCTCCAAAGCGGATCGCGCGCTCGCGCGCGCGCGCGCG 625
QY 424 ProGluProGlnValThrProPheThrGluGluHisProThrHisThr----- 440
Db 624 CCGAGC-----TGGGCGAGCGGCTGGAAA 601
QY 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgVal 460
Db 600 CAGCGCGTCAAGGACCGCTACAGCAC----- 574
QY 461 IleValAspLeuArgTrpPheGly-AlaThrAspProGluAlaAsnAsnLeuValPh 480
Db 573 ACCGTGTCGATGGAGCGCCACGCGCAGCAACATGACCTACCGCGACAAATAG-TCGGACAT 515
QY 480 eGlnAsnAspValGlnAspGlyTyrSerMetPro-----GlnProThrPheArgTyr-- 497
Db 514 CGACCCGACCTACAAAGGACGCTTACGGCCAGCCACTGTTGGCGCATGACCTTCGACTGGAA 455
QY 498 -----ArgProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAs 513
Db 454 CGCAACAGATCCGCATGACCGCTACCGAGCACATGCGCAAGATCGCGCAAGC 395
QY 513 pMetCysGluValAla-----SerAsnLeuGlyGlyTyrLeuProThrSe 528
Db 394 GATGAACCGGAGGCCATCTCGGTGAGCGTCAAGAACTTCGGCGACCACTCAATACAGC 335
QY 528 rProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrArgIleG1 548
Db 334 GGTCTACACG-----ACCACCCACCTGCTCGCGCGGCGATCATGGG 293
QY 548 yPheAspLysAlaThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLe 568
Db 292 CAGCGATCCGAGACCGCGCTCAACCGCTACTCGCAGAGCTGGGACGTGCACAAAGT 233
QY 568 uTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSe 588
Db 232 CTTCGTATGCGCGCTCGGCTTCCCGCAGGCGCACCGCTACACCCCACTGGGCTGGT 173
QY 588 rMetCysHisAlaIleLysSerAlaArgSerIle 599
Db 172 GCGCGCGTGGCGCTACTGCTCGGCGCAAGGCGATC 139
```

RESULT 9

```
US-08-311-731A-131/C
; Sequence 131, Application US/08311731A
; Patent No. 6583286
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 131:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-131
```

```
Alignment Scores:
Pred. No.: 0.125 Length: 42325
Score: 128.50 Matches: 133
Percent Similarity: 31.90% Conservative: 82
Best Local Similarity: 19.73% Mismatches: 222
Query Match: 3.91% Indels: 238
DB: 4 Gaps: 28

US-09-856-327-2 (1-618) x US-08-311-731A-131 (1-42325)

QY 22 IleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAla 41
Db 30466 GTCCCTAAG---GCGGGTCATACCTGGGCGGATGAAGCGGGATTATGACGTCTTAATCATC 30410
QY 42 GlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArgVal 61
Db 30409 GGTTCAGGGTTGGCGGCGATGTCAGTGGCTGCGGCTCACGAAAAAGGTTACCGGGTC 30350
QY 62 ValMetValGluIleGlyAla-----AlaAspSerPheTyrAlaValAlaAsnAlaGlu 79
Db 30349 GGTGTTTGGAGCGGCGCGCTTCGCCGACGAGGATTTTGGCAAGACGTCCTGGGAT 30290
QY 80 GlyThrAlaValProTyrValPro-----GlyTyrHisLysLysAsnGluIleGluPhe 97
Db 30289 CTGCGTAAATTCCTCTGGCGCGCGAAGCTGGGCTGCTACGGCATCAACGCAATTCACCTG 30230
QY 98 GlnLysAspIleAspArgPheValAlaValIleLysGlyAlaLeuGlnGlnValSerVal 117
Db 30229 CTGCGCAAC-----GTGATGATATTGGCCGCGGCC----- 30200
QY 118 ProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProGly 137
Db 30200 ----- 30200
QY 138 SerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAla 157
Db 30200 ----- 30200
QY 158 GluAlaValThrArgGlyValGlyGlyMetSerThrHisThrThrCysSerThrProArg 177
Db 30199 -----GGAGTGGGAGGCGGTCTGTTGAACACTACGCGAATACG----- 30164
QY 178 IleHisProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspPro 197
Db 30163 TTGTACGTTCCCGCGAG-----CCGTTCTTTGCCAACACGACGAG 30125
QY 198 AlaGluAspAspLysGluTrp---AsnGluLeu-----TyrSerGluAlaGluArg 213
Db 30124 TGGGCGCACATCACCGACTGGCAGCAGCGAGCTGGCGCCCGCATACGACCGCGCGCGG 30065
QY 214 LeuIleGlyThr-----SerThrLysGluPheAspGluSer 225
Db 30064 ATGTAGGTGTGGTTTGTAAACCGACCTTCACCGACGCGCGGATACTCAAGAGGTT 30005
```



```
Db 336 CCTGATCAACTGGACC-AGCAGCTCCGCCACCCCGGAA-----CCGACCCTGC 382
Qy 156 erAlaGluAlaValThrArgGlyValGlyGly-----MetSerThrHist 171
Db 383 AGCACTGGCCCGCCAGGTCTACGGGTCAAGGCCACAGCGCGGAGACATGGCGCCCTGGT 442
Qy 171 rpThrCysSerThrProArgIleHisProMetGluSerLeuProGlyIleGlyArgp 191
Db 443 TCGAGAAATGGAGCAGCGCTGCACGTGGCGCCCTGGCGCTCCG-----C 490
Qy 191 roLysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluA 211
Db 491 CGAATGCCAACACGAC-----507
Qy 211 laGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuV 231
Db 508 -----G 508
Qy 231 alLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProL 251
Db 509 TGATCCGC-----C 517
Qy 251 euAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnL 271
Db 518 TGGCTCGCAGAGCTCGC-----TATCACTGGAAGGTCAATCCCGCGCA 562
Qy 271 euPheHisSerIleTyrAsn-----AspAspLysG 281
Db 563 ACGTCCCTGGTGTGGAACCTCGGTACTCGGCATGGGTGCGCGTCAACGCCAAGC 622
Qy 281 lnLysLysLeuPheThr-----LeuLeuThrA 290
Db 623 AGTCGATGCTGTCACCAACATCCCGCGACCCCTCGACAGGTTGGCGAGTGTCTTACC 682
Qy 290 snHisArgCysThrArgLeuAlaLeuThrGly-----G 301
Db 683 TAGCGCGCGCAACCGCTGCTGTGGACGCGACAAAGGTACGGGCTCGAATGCTCGT 742
Qy 301 lyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 321
Db 743 GCATGGATGACGCTGGTGGCGCCCAACGCGCCCGGATC-----783
Qy 321 erSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleG 341
Db 784 -----AGGTCCGCGCGCAGGCACTACGTCTCTCCGCGCGCGCATCA 826
Qy 341 lyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsnA 361
Db 827 ACACCCGCGGATTTCTTCGCTCGAAGCGCCTGATCC-GAGCCACGCGGTTCGCAAG 885
Qy 361 spSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGlnI 381
Db 886 CGCACCTT---CCTGCACAGGTGAATTCAGCGGGGCTGTTCCGACCGGGTATCAAC 942
Qy 381 leValLeuArgGlnGluPheValAspSerValArgAspProTyrGlyLeuProTrpT 401
Db 943 CCCTTCTACGGCGCACCGCAGTCGATCTACTCCGACCATTTCCCAATGGAGCGGTC 1002
Qy 401 rpLysGluAlaValAlaGln-----407
Db 1003 ACCGGACGCATCTCTCAAACTGGAGTGGCGCGCTGCAACGCTGCTGGCTCGGTG 1062
Qy 408 -----HisIleAlaLysAsnProThr-----414
Db 1063 CTGCTCGGCGCTTCGGCAGGACACGCCCTGCGCATGGAAACAGCTACCGCATACCAAC 1122
Qy 415 --AspAlaLeuProIleProPheArgAspProGluProGln-----ValThrThrP 431
Db 1123 ATGATGTCGCGCTGCTCGCGACGGCTTCACCCCGGACAGCGGAGGGCTCGGTCCGC 1182
Qy 431 ropPheThrGluGlnHisProThrHisThrGlnIleHisArgAspAlaPheSerTyrGlyA 451
Db 1183 CTGCGCGCGCAGCACACCC---AGTGTCTGACTACCGGATGACCGACTACACCTGGGAC 1239
```

```
Qy 451 laValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThrA 471
Db 1240 GGTATCCGCGCTGCTTCCA-----1259
Qy 471 spProGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetP 491
Db 1260 -----TACCATGGCG 1269
Qy 491 roGlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetM 511
Db 1270 GAGATCCAGTTCGCGCGCGCGGAACTGCGCGGTCACAGCGACGCGCGGTAGC 1329
Qy 511 etAlaAspMetCysGluValAlaSerAsn-----LeuGlyGlyTyrLeuProT 527
Db 1330 CGGAAGAACCTGCGAGGAGCGCGGATCATCGACGCGCTGCGCTGGAGATCTACGCG 1389
Qy 527 hrSerProGlnPheMetAspProGlyLeuAlaLeu-HisLeuAlaGlyThrThrArg 546
Db 1390 ACGCGCTGGGCAG-----CGCCACGTATGGCGGCTGCGCC 1428
Qy 547 IleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAla 566
Db 1429 ATGGGCGAAGACCGCGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1488
Qy 567 AsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeu 586
Db 1489 AACCTGTGATCCACGACGCGTCTGTTCCACACCATCGGCGGCGGCGGCGGCGG 1548
Qy 587 Thr 587
Db 1549 TCG 1551
```

RESULT 11
US-09-265-108-1
; Sequence 1, Application US/09265108
; Patent No. 6033891
; GENERAL INFORMATION:-
; APPLICANT: Golightly, Elizabeth
; APPLICANT: Brown, Kimberly
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: Having Cellobiose Dehydrogenase Activity
; CURRENT APPLICATION NUMBER: US/09/265,108
; CURRENT FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2576
; TYPE: DNA
; ORGANISM: Humicola
US-09-265-108-1

Alignment Scores:
Pred. No.: 0.00366 Length: 2576
Score: 124.00 Matches: 131
Percent Similarity: 34.70% Conservative: 89
Best Local Similarity: 20.66% Mismatches: 270
Query Match: 3.78% Indels: 145
DB: 3 Gaps: 30

US-09-856-327-2 (1-618) x US-09-265-108-1 (1-2576)

```
Qy 26 AlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyPro 45
Db 794 GCCCTGAGGAGACTTAT-----GACTATATTTGTTGTTGGTGGCGGCC 838
Qy 46 IleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArgValValMetValGlu 65
Db 839 GCGCGTATTCCTGTCGCGGCAAGCTGAGCGAGCGCGGCGGCGGCGGCGGCGGCGG 898
Qy 66 IleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGlyThrAlaValPro--- 84
Db -----
```



```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11428
; LENGTH: 2106
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11428

```

Alignment Scores:		
Pred. No.:	0.00382	2106
Score:	122.50	Matches: 130
Percent Similarity:	34.95%	Conservative: 65
Best Local Similarity:	23.30%	Mismatches: 232
Query Match:	3.73%	Indels: 142
DB:	4	Gaps: 30

US-09-856-327-2 (1-618) x US-09-252-991A-11428 (1-2106)

26	Ala	Ile	His	Glu	Thr	Tyr	Gly	Asn	Asp	Gly	Val	Asp	Val	Phe	Ile	Ala	Gly	Ser	Gly	Pro	45
385	CGC	ATC	CA	CC	CGC	CGC	TAC	GGC	---	---	---	---	---	---	---	---	---	---	---	---	420
46	Ile	Gly	Ala	Thr	Tyr	Ala	Lys	Leu	Cys	Val	Glu	Ala	Ac	Lys	Leu	Arg	Val	Met	Val	Glu	65
421	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	474
66	Ile	Gly	Ala	Ala	Asp	Ser	Phe	Tyr	Ala	Val	Asn	Ala	Glu	---	---	---	---	---	---	---	81
475	GCG	GGC	GC	ATC	AGC	CC	ATG	GC	CA	AGT	CGC	CC	CA	AGC	GC	GT	GAT	G	AA	GA	534
82	Ala	Val	Pro	Tyr	Val	Pro	Gly	Tyr	His	---	---	---	---	---	---	---	---	---	---	---	99
535	GCG	TG	CGC	TG	TG	TG	CGC	GT	AT	CA	CGC	GA	AGC	CC	CA	AGC	CT	C	G	A	594
100	Asp	Ile	Asp	Arg	---	---	---	Phe	Val	Asn	Val	Ile	Lys	Gly	Ala	---	---	---	---	---	117
595	GA	CCG	GGC	ATC	GC	TAC	CGG	TG	CT	CT	CA	AGC	CC	CGC	GT	GGC	GGC	GGC	GGC	GA	654
117	Pro	Val	Arg	Asn	Gln	Asn	Val	---	---	---	---	---	---	---	---	---	---	---	---	---	131
655	GGC	AT	GA	AGT	GGT	TGG	AGC	CGC	AGC	CGC	AGC	GT	GGC	AGC	GC	CTT	TCC	TCC	CC	CA	714
131	atr	P	Ser	Ala	Pro	Pro	Gly	Ser	Ser	---	---	---	---	---	---	---	---	---	---	---	151
715	GAG	CGC	AAG	CGC	CGC	TTC	GGC	AGC	CGC	GT	ATC	GT	G	T	G	T	G	A	A	A	764
151	lu	Phe	Glu	Asn	Leu	Ser	Ala	Ala	Val	Thr	Arg										

Qy	266	HisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPhe	285
Db	1153	CACCCATCGAA-----GTACGGCTCTATCGGAAGACCCCGAGCGGACGTTCCCT	1203
Qy	286	ThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly---GlyTyrGluLys	304
Db	1204	CCCCCAGCGCGCCGTGATGCTATACCGGAAGCGCGCGCGCCGCCCGCGCGGGGT	1263
Qy	305	LysIleGlyAlaAlaGlu-----ValArgAsn	313
Db	1264	GACAGCGGGTACGCGAAGCGCAGAGGTCTCGCGGTCTACGACCCGATGCTGCCCAAG	1323
Qy	314	LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLysValTyr	333
Db	1324	CTGATCGCCCTGGCGGAACCCCGCGAAGACCGCCGACGCGCTGCTGGCG-----	1374
Qy	334	ValLeuAlaSerGlyAlaIleGlyAsn-----ProGlnIle	345
Db	1375	ATGCTGGCGAAACCTCGTTCGGCGGCTCGGCACCACTGGCGTTCCTCCGCCGATC	1434
Qy	346	LeuTyrAsnSerGlyPheSerGlyLeuGln-----ValThrProArgAsn	360
Db	1435	CTCGGCATCCGGGTTTCGCGCGCGCGAACTGGACACCGGGTTCATCGCGCCACCAA	1494
Qy	361	AspSerIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln	380
Db	1495	GAGCACCTGTCTGGCGCGCGCGAAGCGCTCGCGGAGCATTTCTGGCAAGCCGCCGAA	1554
Qy	381	IleValLeuArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProTyr	400
Db	1555	GCCTGGCTGAGAGCAACCGCGCCACCGCTCGCGCAGCAGCCCGCATTCG---CGGTGG	1611
Qy	401	-----TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAsp	415
Db	1612	AGCGCAACGACGCGTGGCGACGGCGCTCGCGCGGA-GTC-----CGACCTGATGCT	1664
Qy	416	AlaLeuPro-----IleProPheArgAspProGluProGlnValThrThrProPheThr	433
Db	1665	GCCTCGCGCAGCAGCCCGCTGCTGCGCCCTCGCGCCAGCCGACCCAGCAGATCG	1724
Qy	434	GluGlu-----HisProTyrHisThrGlnIle	442
Db	1725	CCTCGAGCGGACGACTGTGTCAGCGGGGTGGACGGTGTCAACCG-----	1769
Qy	443	HisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgValIleVal	462
Db	1770	-----TCGCTCGCGCGGTTCGCGCG-----CGAGCGCCAATT	1802
Qy	463	-AspLeuArgTyrPheGly-----AlaThrAspProGluAlaAsnAs	476
Db	1803	GTTCCTCGAATGGAGGGCAACTGCTGGCCATCGAAGCGCTGACCCGATGCCGAGGC	1862
Qy	476	nLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetPro	491
Db	1863	CGAGCGCGCCATGCGCCAC-----CAGGCGGCTTGAGCGCGCGC	1902

```

RESULT 14
US-08-930-001-1
; Sequence 1, Application US/089300001A
; Patent No. 6281412
; GENERAL INFORMATION:
; APPLICANT: MURATA, No. 6281412iio
; TITLE OF INVENTION: METHOD FOR PRODUCING OSMOTOLERANT PLANTS
; FILE REFERENCE: 0230-118P
; CURRENT APPLICATION NUMBER: US/08/930, 001A
; CURRENT FILING DATE: 1997-09-26
; EARLIER APPLICATION NUMBER: JAPAN 106819/1995
; EARLIER FILING DATE: 1995-03-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2400

```

```

; TYPE: DNA
; ORGANISM: Arthrobacter globiformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (361)..(2001)
US-08-930-001-1

```

Alignment Scores:

pred. No. : 3	0.00475	Length:	2400
Score:	122.50	Matches:	138
Percent Similarity:	31.90%	Conservative:	70
Best Local Similarity:	21.17%	Mismatches:	258
Query Match:	3.73%	Indels:	189
DB:	3	Gaps:	30

US-09-856-327-2 (1-618) X US-08-930-001-1 (1-2400)

QY	20	GlyGlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyVal-AspValph	39
Db	351	GGAGAACTAGATGCATGCACAACTGAGCAACCTGAGCAGAGGATTTCGACTACAT	410
QY	39	eileAlaGlySerGlyProIleGlyAlaThrTyr--AlaLysLeuCysValGluAlaGl	58
Db	411	CGTCGTCGGCGGGTCCGCGGGGGCGCGTCGCCCGCGTGCAGCAGGATTCGCCG	470
QY	58	yLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGl	78
Db	471	AGTGAGCGTGGCGTGTGGAGGCGGCCCGGATGAC-----	507
QY	78	uGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGluPheGl	98
Db	508	-----CGCGGCGTGCCTCCGAGGTG-----	525
QY	98	nLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSerValPr	118
Db	526	-CTGCACCTGGACCGCTGGATGGAGCTGCTGGAAATCGGGTACGACTGG---GACTACCC	581
QY	118	oValArgAsnGln-AsnValProThrLeuAsp-----ProGlyAlatrps	133
Db	582	GATCGAGCCGCGAGGAGAACGCAACTCCTTCATCGCCCATGCCCGTGCACAGGTTCATGGG	641
QY	133	erAlaProProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheG	153
Db	642	CGCTGCTCCAGCCCAACTCTGTCATCTGCCTCT-----	676
QY	153	luAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrC	173
Db	677	-----GGCCCGCGCGGACCTTGACGA701	701
QY	173	ysSerThrProArgIleHisProProMetGluSerLeuProGlyIleGlyArgProLysL	193
Db	702	GTGGGAGGCCAAGTACGCGCCACCG-----GCTGGACGCGCGAGG-742	742
QY	193	euSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluAla---	211
Db	743	-----CGGCTGGCCGCTGTACAGCGCTGGAACCAACAGGACGCGGCGCCCGGA794	794
QY	212	--GluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuV	231
Db	795	CGCGCGCGCACCGGGACTCCCGCC-----CCGTGCACCTGATGAACGTGCC842	842
QY	231	alleuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProL	251
Db	843	CCCGAAGGACC-----CGACCGGCGTCGCGCTCTGTGACGCCCTCGCAGCA887	887
QY	251	euAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTyrHisSerAlaGluAsnL	271
Db	888	GGCGCGCATCCCGCGCGAAGTTCACACCGGCACCCGCTGTCGTAACGGCGCC-AACT946	946
QY	271	euPheHis-----SerIleTyrA77	77
Db	947	TCCTTCAGATCAACCGGCGCGGAGCGGACCGGCTCCTCAGCTCGGTCTCTCTACATCC1006	1006

Qy	277	snAspAspLysGlnLysLysLeuPheThrLeuLeuLeuThrAsnHisArgCysThrArgLeuA	297
Db	1007	ACCGCATCGTGCAGCAGGAAACTTACCCCTGCTTAACCGGCTCGCGCGCCGACAGCTGG	1066
Qy	297	laLeuThrGly-----GlyTyrGluLysLysIleGlyAlaAlaGluValA	312
Db	1067	TGTTTCGACGGGACAGCGGTGTACCGCGTGTGCACATCGTGAGACTCGCCTTCGGCGCGCA	1126
Qy	312	rgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuLeuAspSerTyrIleMetalalysv	332
Db	1127	CCCATCGCTCGCGCGCGCAATGAA-----	1152
Qy	332	alTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheS	352
Db	1153	-GTGCTGCTCTCCACCGCGCGCATCATACGCGGAAGCTGTTGATCTCTCGGAATCG	1210
Qy	352	erGlyLeu-GlnValThrPro-----ArgAsnAspSerLeuIle-----ProAsnLeu	367
Db	1211	GCCCCGCGCCCACTCGCCGACGACCGCATCGAGTCCCTGTGTGGACTCCCCCGCGGTG	1270
Qy	368	GlyArgTyrIleThrGlnGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPhe	387
Db	1271	GGCAGCACTGAGGACCCCGGAAGCGGTGTCAGTTCGAGGCCACAGAGCCCATG	1330
Qy	388	ValAspSerValArgAspAspProTyrGlyLeuProThrTrpLysGluAlaVal-----	405
Db	1331	GTCCCGCAG-----TCCACGCACTGTTGGGAGATCGGCATCTTCACCTC	1372
Qy	406	-----AlaGlnHisIleAlaLysAsnProThr	414
Db	1373	CCCACCGAGACGGCTGTGGACCGCCCGACCTGATGATGCATACCGCTCCGTGCCGTTC	1432
Qy	415	AspAlaLeuProfileProPheArgAspProGluProGlnValThrThr-----	430
Db	1433	GACATGAACACCTCGCGCAGCTACCC-----ACCAGGAGACGGGCTT	1480
Qy	431	---ProPheThrGluGluHis-----ProThrPheThrGlnIleHisArgAsp	445
Db	1481	CAGCCTCACCCGAACTCATCGACGCGCCGCTCCCGCGCACTGTCG-----	1528
Qy	446	AlaPheSerTyrGlyAlaValGlyProGluValAsp-SerArgValIleValAspLeuAr	465
Db	1529	-----GCTGCGACCGCGACTTCCCGGATAAGCCCATGTGTCGACCGCG	1573
Qy	465	qTrrPheGlyAlaThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsn-----As	483
Db	1574	CTACTTC-----ACCGACCCAGAGGGCC-ATGACATCGCGCTCATGTGTCGCGCATC	1626
Qy	483	pValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlase	503
Db	1627	CGAAAGCCCGCGAAATCGCGCCGCGCCATGCGCGAATGGACGCGCGCAGCTC	1686
Qy	503	rAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGlyGl	523
Db	1687	TCCCCCGCGCTCGAGCGCGACCGCAGGAGCTGCAGGACTACATCC-----	1735
Qy	523	yTyrLeuThrSerProProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGl	543
Db	1736	-GCAAGACGCACACACCG-----TCTAC-CACCCCGTGGG	1769
Qy	543	yThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTr	563
Db	1770	CACCGTCGCATGGC-----GCGGTGAGGACGAGATGTCCCGCGTC-----	1812
Qy	563	pAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGlu--	582
Db	1813	-----GACCCCGAGCTCGGGTCAAGGCGTCAACCGGTCTCGCGTCTCGCGCGCTCGGT	1868
Qy	583	-----AsnProThrLeuThrSerMet-----	589
Db	1869	CATGCCCGACAGCTGACCGTCAACCCACATCAACCGTCATGATCATCGGCGAGCGCTG	1928
Qy	590	-CysHisAlaIleLysSerAlaArgSer	598

Db 1929 CGCGGACCTTATCGCTCCGCGCGCGC 1956
 |||:::|||||||
 RESULT 15
 US-09-252-991A-11409
 ; Sequence 11409, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 11409
 ; LENGTH: 3009
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-11409

Alignment Scores:
 Pred. No.: 0.00692 Length: 3009
 Score: 122.50 Matches: 130
 Percent Similarity: 34.95% Conservative: 65
 Best Local Similarity: 23.30% Mismatches: 222
 Query Match: 3.73% Indels: 142
 DB: 4 Gaps: 30

US-09-856-327-2 (1-618) x US-09-252-991A-11409 (1-3009)

Db	1929	CGCGGACCTTATCGCTCCGCGCGCGC	1956
Qy	26	AlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyPro	45
Db	365	CGCATCCACCCGCGCTACGGC	400
Qy	46	IleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArgValValMetValGlu	65
Db	401	-----GCCGACTTCGCGCGCGCTCGAGGAAGCGCGCTCTCTCTCGGAGAAC	454
Qy	66	IleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGlu	81
Db	455	CGCGCGCCATCGACCCATATGGCCACCAAGTCGGCCGCCAAGCGCTGATGGAAGAAGCC	514
Qy	82	AlaValProTyrValProGlyTyrHis	99
Db	515	GGCGTCCGCTGTGTGCGCGCTATCACGCGGAAGCCAGACCTTCGAGACTTTCGCGCGG	574
Qy	100	AspIleAspArg	117
Db	575	GAGCGCGCGCATCGCTACCGGTGCTGCTCAAGCGCGCGCTGGCGCGCGCGCAAG	634
Qy	117	lProValArgAsnGlnAsnVal	131
Db	635	GGCATGAAGTGTGGAGCGCGAGCGCGAGCTGGCGAGCGCTTTCCTCCGCGCAGCGC	694
Qy	131	aTrpSerAlaProProGlySerSer	151
Db	695	GAGGCCAAGCGCGCTTCGCGGACGCGCTATGCTGGTGGAGAAATACCT	744
Qy	151	luPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMet	170
Db	745	-----GCTGAAGCGCGCGCCACGTCGAGATCCAGGTATTCGCCCGCAGCGC	787
Qy	170	IsTrpThrCysSerThrProArg	184
Db	788	CATGCCCACTGCCTACTCTAACCAACGCGACTGTTCCATCCAGCGCGCGCCACCAAG	847
Qy	184	erLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAspLysGluT	204
Db	848	GTCTCTCAGGAAGCCCGCGCGCTCTCGCGCGGAAGTCCGCGCGGCATCGGCGAG	907
Qy	204	rpAsnGluLeuTyrSerGluAlaGluArgLeu	222
Db	908	GCGCGGTACGCGCGCGCGAGCGATCGGTACGTCCGCGCGGTACCGTTCCTC	967
Qy	223	-----AspGlu	228
Db	968	CTCAGCAAGCGCGCGCGAGTCTTCTTCATGAAATGAACACCCCGCTCGAGTGAACAC	1027
Qy	229	-----ThrLeuValLeuArgSerLeuGln	245
Db	1028	CCGCTACCGAGGCGATCACCGCTCGACCTGCTCGCGATGCGCGTCCGCTCGC	1087
Qy	246	PheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTyr	265
Db	1088	GGCGAGCGCTACCGCTGACGCGAGGACAGTACCG	1132
Qy	266	HisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPhe	285
Db	1133	CAGCGCATCGAA	1183
Qy	286	ThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly	304
Db	1184	CCGCGCAGCGCGCTGATGCTATACCGGAAGCGCGCGCGCGCGCGCGGTC	1243
Qy	305	LysIleGlyAlaAlaGlu	313
Db	1244	GACAGCGGGTACGCGAAGCGGAGGCTCCCGCTTCTACACCGGATCGTGGCGCAAG	1303
Qy	314	LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetalLysValTyr	333
Db	1304	CTGATCGCTGGGCGGAACCCCGAGGAAGCCCGCAGCGCTGCTGGCG	1354
Qy	334	ValLeuAlaSerGlyAlaIleGlyAsn	345
Db	1355	ATGCTGGCGAAACCTCGGTGCGCGCGCTCGCGCAACCTGCGTTCCTCCCGCGCATC	1414
Qy	346	LeuTyrAsnSerGlyPheSerGlyLeuGln	360
Db	1415	CTCGGCATCCGCGCGTTCGCGCGCGGACGACACCGGGTTCATCGCGCGCCACCA	1474
Qy	361	AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln	380
Db	1475	GACGACTCTGCGCGCGCGCAAGCGCTGCGGAGCATTTCTGGCAAGCGCGCGCGAA	1534
Qy	381	IleValLeuArgGlnGluPheValAspSerValArgAspProTyrGlyLeuProTyr	400
Db	1535	GCTGCTCGAGCGAAGCGCGCGCACCTCGCGGAGCAGCCGCGCATTCG	1591
Qy	401	-----TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAsp	415
Db	1592	AGCGCAACGAGCGCTGGCGAGCGCGCTGGCGCGGA-GTC	1644
Qy	416	AlaLeuPro	433
Db	1645	GCCTGCGCGCAGCAACCGCTGCTGCGCTGCGCGCACCGCGCGCAGCATTCG	1704
Qy	434	GluGlu	442
Db	1705	CCTCAGCGCGCAGCGCTGCTGCGCGGTGGAGCGGTGTCACCGC	1749
Qy	443	HisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgValIleVal	462
Db	1750	-----TCGCTCGCGCGGTGGCGCG	1782
Qy	463	AspLeuArgTrpPheGly	476
Db	1783	GTTCCTCGAATGGGAGGCGGAACCTGCTGGCCATCGAAGCGCGCGCATTCGCGAGCG	1842
Qy	476	nLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetPro	491
Db	1843	CGAGCGCGCGCATGCCAC	1882

Search completed: September 24, 2003, 10:58:43
Job time : 7718 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 05:33:29 ; Search time 3509 Seconds
(without alignments)
4280.469 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSTEQMLRDPYRSMQING.....IINTLKGTGDKNGTGEHRNL 618

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n_model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09856327/runat_23092003_153419_15517/app_query.fasta_1.775
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09856327.ecgn.1.1.2810.runat_23092003_153419_15517 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_fod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gssl.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	241	7.3	449	9	AI213008
2	120.5	3.7	1971	13	BU227568 y6e01al.r
c 3	117.5	3.6	384	29	CJE315007 Campyloba
4	112.5	3.4	1145	12	BM462502 AGENCOURT
5	112	3.4	834	10	BG541504 602570950
6	112	3.4	839	28	BH451002 BOGXY12TR
7	109	3.3	1451	11	AK029462 Mus muscu
8	109	3.3	2996	11	BC011088 Mus muscu
9	107	3.3	690	12	BJ258831 BJ258831
10	107	3.3	1319	13	BU518146 AGENCOURT
c 11	106.5	3.2	3111	11	AK085667 Mus muscu
12	106	3.2	691	9	AV764246 AV764246
c 13	106	3.2	783	28	BH494670 BOGVJ59TR
c 14	106	3.2	920	10	BE782505 601467649
15	106	3.2	4866	11	BC034946 Homo sapi
16	105.5	3.2	605	12	BM425539 pgf2c.pk0
17	105.5	3.2	1201	9	AL536149 AL536149
18	105.5	3.2	1288	13	BU541981 AGENCOURT
19	104	3.2	2518	11	BC003719 Mus muscu
20	103.5	3.2	976	13	BQ713154 AGENCOURT
c 21	103.5	3.2	988	11	CNS08JNK BX014892 Single re
c 22	103	3.1	503	9	AW255795 ML863 pep
c 23	103	3.1	661	12	BJ136654 BJ136654
24	103	3.1	1003	13	BU962960 AGENCOURT
25	103	3.1	1819	11	BC031681 Homo sapi
26	103	3.1	1869	11	BC040952 Homo sapi
27	103	3.1	1924	10	BG027879 BG027879
28	102.5	3.1	1023	13	BU124024 603147663
29	102.5	3.1	1514	11	BC051546 Mus muscu
30	102.5	3.1	2910	11	AK078478 Mus muscu
31	102	3.1	633	9	AL957104 AL957104
c 32	102	3.1	795	28	BH708030 BOMDQ12TF
33	102	3.1	886	28	BH972564 odh12a06
34	102	3.1	3611	11	AK053309 Mus muscu
35	101.5	3.1	766	12	BM291983 EST574525
36	101.5	3.1	1201	13	BX423637 BX423637
37	101	3.1	707	12	BM609508 170006871
c 38	101	3.1	733	28	BH524476 BOGOK35TF
39	101	3.1	822	10	BF500573 AT15559.5
c 40	101	3.1	905	10	BE616603 601278816
41	101	3.1	977	29	CC357805 PURPH38TD
42	100.5	3.1	732	28	BH601509 BOGSEB77TF
43	100.5	3.1	1031	12	BM469849 AGENCOURT
44	100.5	3.1	1230	29	CC187722 CH261-26B
45	100.5	3.1	2592	11	AK045276 Mus muscu

ALIGNMENTS

RESULT 1
AI213008
LOCUS y6e01al.r1 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Emericella nidulans cDNA clone y6e01al 5', mRNA sequence.
DEFINITION AI213008
ACCESSION AI213008
VERSION AI213008
KEYWORDS EST
SOURCE Emericella nidulans (anamorph: Aspergillus nidulans)
ORGANISM Emericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; Emericella.

```

REFERENCE 1 (bases 1 to 449)
AUTHORS  Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
        Prade,R. and Roe,B.
TITLE     An Aspergillus nidulans EST Database
JOURNAL   Unpublished
COMMENT   Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
        Department of Chemistry and Biochemistry
        Advanced Center for Genome Technology, University of Oklahoma
        620 Parrington Oval, Norman, OK 73019, USA
        Tel: 405 325 4912
        Fax: 405 325 7762
        Email: broe@ou.edu
        We anticipate the future release of the cDNA clones to the Fungal
        Genetics Stock Center
        Seq primer: T3
        High quality sequence stop: 442.
        Location/Qualifiers
            1..449
                /organism="Emmericella nidulans"
                /mol_type="mRNA"
                /strain="FGSC A26"
                /db_xref="taxon:162425"
                /clone="y6e01a1"
                /tissue_type="vegetative mycelia, asexual structures"
                /clone_lib="Aspergillus nidulans 24hr asexual
        developmental and vegetative cDNA lambda zap library"
                /notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
        XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
        3' end of cDNA cloned into XhoI site of pBluescript"
        BASE COUNT. 130 a 108 c 115 g 96 t
        ORIGIN

Alignment Scores:
Pred. No.:      3.05e-13      Length:      449
Score:          241.00        Matches:    58
Percent Similarity: 52.90%    Conservative: 15
Best Local Similarity: 42.03% Mismatches:   41
Query Match:     7.34%       Indels:     24
DB:              9           Gaps:         3

US-09-856-327-2 (1-618) x AI213008 (1-449)

Qy  37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  80 GATGTCCTAATCATCGGCTCGGCCCGCCAGTCGGGGCAACTTACCGAAGGAGATCTGGAC 139
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  57 AlaGlyLeu-----ArgValValMetValGluIleGly 67
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  140 CCCGGCTCCGGTGCTTCGCCTGGAGAAAAGCTCCAAAGGTTATTATGTGGAAACTGGG 199
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  68 AlaAlaAspSerPheTyrAlaValAsnAlaGluGlyThrAlaValProTyrValPro 87
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  200 GCCACAGAGTCT-----AAAGTGCCT 220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  88 GlyTyrHisLysLysAsnGluIleGluPheGlnLysAspIleAspArgPheValAsnVal 107
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  221 GGTGAACACAGAAGAAATGCCGTGGTATACCAACAACACATGATCTCTTTGCAATGTT 280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  108 IleLysGlyAlaLeuGlnValSerValProValIargAsnGlnAsnValProThrLeu 127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  281 ATTCAAGGGAGCGCTATTTCGAACCTCTGTCCCAACACAGCTGTGGATCCGAACCTCAAGTTG 340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  128 AspProGlyAlaTrpSerAlaProProGlySerSerAlaIleSerAsnGlyLysAsnPro 147
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  341 CTTCCCGTCTCTTGGTCT-----CCACGTGAAAAACAACAACTTCAACGGCGCAACAAG 394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  148 HisGlnArgGluPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGly 165
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  395 GACCAGAATATCTACCACAACTTAGACGCTAATGCTGTGTCGCCCAATGTTGGC 448
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
BU227568 linear EST 26-NOV-2002
LOCUS

```

```

DEFINITION 603947501F1 CSEQCHN23 Gallus gallus cDNA clone ChEST902b2 5', mRNA
sequence.
ACCESSION BU227568
VERSION BU227568.1 GI:25465582
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1971)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST
)
PO Box 88, Manchester, M60 1OD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES             source
    1..1971
        /organism="Gallus gallus"
        /mol_type="mRNA"
        /strain="White Leghorn, Hixex"
        /db_xref="taxon:9031"
        /clone="ChEST902b2"
        /dev_stage="22"
        /lab_host="DH10B"
        /clone_lib="CSEQCHN23"
        /note="Organ: heads; Vector: pBluescript II KS(+); Site:1:
        ECORI; Site 2: NotI; This normalized library was
        constructed from 1 million independent clones. cDNA
        synthesis was initiated using an oligo(dT) primer, using
        methylated C in the first strand synthesis reaction.
        Following this first strand reaction, double-stranded cDNA
        was bluntended, ligated to NotI adapters, digested with ECORI
        , size-selected, and cloned into the NotI and ECORI
        compatible sites of a custom modified MCS of the
        pBluescript (KS+) vector. The library was normalized in 2
        rounds using conditions adapted from Soares et al., PNAS
        (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
        (1996): 791, except that a significantly longer
        reannealing hybridization was used."
        BASE COUNT 564 a 847 c 370 g 190 t
    ORIGIN
Alignment Scores:
Pred. No.:      1.04      Length:      1971
Score:          120.50    Matches:      79
Percent Similarity: 33.74%  Conservative: 59
Best Local Similarity: 19.32% Mismatches:    180
Query Match:      3.67%   Indels:      91
DB:               13      Gaps:      12

US-09-856-327-2 (1-618) x BU227568 (1-1971)

Qy 116 SerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaPro 135
   ::: ||| ::: ||||| ||||| ::: |||
Db 755 AACACCCCAACACCGGCCCAACACCGGCCCAACCTCGCGACCTGGCCCAACGCCCCC 814
   ::|||
Qy 136 ProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeu 155
   |||| ::|||::: :::: |||
Db 815 CAGGGCAGCAGCGGCCCAACACCGGCCCAACCCCTACACCCCAACGCGCGCACACAC 874
   ::::

Qy 156 SerAlaGluAlaValThrArgGlyVal-----GlyGlyMetSerThrIstIrpThrCys 173
   :::: ||| |||||::: ::::
Db 875 AACCCCCCGCAAAACACAAAGACCTCGGGCGCAACCTCGCGCGCACACACGGCTCTTCT 934
   ::::

```

```

Qy 174 SerThrProArgIleHisProPrometGluSerLeuProGlyIleGlyArgProLys--- 192
Db 935 CCACGACCCATCAAGTTCCACCCCAAAAGACACCCGCCCATCAATCCCCCAACCA 994
Qy 193 -----LeuSerAsnAspProAlaGluAspLysGluTrpAsnGluLeuTyrSerGlu 210
Db 995 TTCCACTATGTGCCCCCCATTTTCACCAACGACGCGCCACAGAGCGGTCAATCAAA 1054
Qy 211 AlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeu 230
Db 1055 GCACCCCGGACCAACACTTCAACAGCGGACGCGCATTCCTCCACACACACCATG 1114
Qy 231 ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgPro-LeuPr 250
Db 1115 GCGCACACCCAGCGCAAAACACAGCAATGCCCGCAACGACCCACACGACCATCCACC 1174
Qy 250 oLeuAlaCysHis-----ArgLeuLysAsnAlaProGluTyr 262
Db 1175 AGCAGGCCCCATCAGCGACCCAGCAACCGCCCAACAGCGGACGCGGCAT 1234
Qy 262 rValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLysGlnLys 282
Db 1235 AAGCGCACACGCGGCAACAAACCTACACACGACGCGCCGCCGCCCAACAGCGAAA 1294
Qy 282 sLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGlyGly-- 301
Db 1295 ACCTACACAGACATATCTCGACGCGCCAGAGCAAAAGG-----GGGGAGC 1342
Qy 302 -----TyrGluLysLysIleGlyAlaLeuAlaGluValAr 312
Db 1343 CGACACCGCGGACGCGCTACTCCACCATATCAGGACCTATACCGCGCGCGGGAC 1402
Qy 312 gAsnLeuLeuAlaThrArgAsnProSerSerGlnLeu----- 324
Db 1403 CAGACCATACACCCACCAACCAAGACCCCAATTTACTCCACCTCATATTTATGGGG 1462
Qy 324 ----- 324
Db 1463 CAGACAACACACTCAGCGCGCTTACACACCATTTTCACAGCAGCACACCCGCGCA 1522
Qy 325 -----AspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAl 339
Db 1523 CGCCCTTATAAAACCACTCTTTCCACCCAGCGGACTTAATAATACATAGCAACGC 1582
Qy 339 aIleGlyAsn-----ProGlnIleLeuTyrAsnSerGlyPheSerG1 353
Db 1583 GCTCAATAATATACACGACAGAAATTTCCGCCCAACACACACAACTTTAACAAATC 1642
Qy 353 yLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrIleThrG1 373
Db 1643 AGCAGATATATCGCCGACAGACAGACAACTACCCAC----- 1682
Qy 373 uGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAs 393
Db 1683 ----CCGCTA---TATACAAACCCAGCGCGCAGAGACAG-----CCAGGTTGGGGGA 1729
Qy 393 pAspProTyrGlyLeuProTrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnPr 413
Db 1730 CAAGCCACACACTC-----CACGAACACACAAAGCAGGTATATCACAAAATCC 1780
Qy 413 oThrAsp-----AlaLeuProIleProPheArgAsp-P 424
Db 1781 CACCGAGNACACTCACCCCGGCACTATACACAGCGCGCCACACCATAGCAGCGGACGA 1840
Qy 424 roGluProGlnValThrProPheThrGluGluHisProTrpHisThrGlnIleHisA 444
Db 1841 CCGCAAGCGCAGCGCAACACCCCATCAAAAGACACCGCGCCCATAGGCGCGAGCACA 1900
Qy 444 rAspAlaPheSerTyrGlyAla 451
Db 1901 CACACCGCGCACACACGCGGCGC 1923

```

```

RESULT 3
CJ315007/c 384 bp DNA linear GSS 16-MAY-2002
LOCUS Campylobacter jejuni GSS, clone 46, genomic survey sequence.
DEFINITION
ACCESSION AJ315007
VERSION AJ315007.1 GI:20975577
KEYWORDS GSS; genome survey sequence.
SOURCE Campylobacter jejuni
ORGANISM Campylobacter jejuni
          Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
          Campylobacteraceae; Campylobacter.
REFERENCE
  1 Ahmed,I.H., Manning,G., Wassenaar,T.M., Cawthraw,S. and Newell,D.G.
    Identification of genetic differences between two Campylobacter
    jejuni strains with different colonization potentials
    Microbiology 148 (Pt 4), 1203-1212 (2002)
    21930301
    11932464
  2 Ahmed,I.H.
    Direct Submission
    Submitted (15-JUN-2001) Ahmed I.H., Department of Bacterial
    Diseases, Veterinary Laboratories Agency, Woodham Lane, New Haw,
    Addlestone, Surrey, KT15 3NB, UNITED KINGDOM
    Location/Qualifiers
      1..384
        /organism="Campylobacter jejuni"
        /mol_type="genomic DNA"
        /strain="81116"
        /db_xref="taxon:197"
        /clone="46"
        /note="obtained by genomic subtractive hybridisation"
BASE COUNT 122 a 62 c 72 g 128 t
ORIGIN
Alignment Scores:
Pred. No.: 0.216 Length: 384
Score: 117.50 Matches: 33
Percent Similarity: 42.02% Conservative: 17
Best Local Similarity: 27.73% Mismatches: 62
Query Match: 3.58% Indels: 7
DB: 29 Gaps: 1
US-09-856-327-2 (1-618) x CJ315007 (1-384)
Qy 476 AsnLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPhe 495
Db 383 AATTATTATCCCTTGATCCCACTTATAAAGATGCTTATGACCTACCACTTATCGCGTA 324
Qy 496 ArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCys 515
Db 323 ACCTATAATTTTACCACAGCAAGATAAATCTTTATAATTTTATGCTTCCAAATAAAA 264
Qy 516 GluValAlaSerAsnLeuGlyTyrLeuProThrSerProGlnPheMetAspPro 535
Db 263 AGCATCATGCAAGAAATGGGAGCGCAAAATATTACACTTCGCCAAAATTAGGAGATTAT 204
Qy 536 GlyLeu-----AlaLeuHisLeuAlaGlyThrArgIleGly 548
Db 203 GGAGTAGTTTCAATGGTGGTATATCTCTACCATATTTGTCGGGGAACATATATGGGT 144
Qy 549 PheAspLysAlaThrThrValAlaAspAsnSerLeuValTrpAspPheAlaAsnLeu 568
Db 143 ACAGATAAAACGACAAAGGTAGTTAACTCTTTATACAGCATTTGGAGTGTAAATCTT 84
Qy 569 TyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThr 587
Db 83 TTTTCGTAGGTGCGAGGAATTTTCCACATAATCCAGGCTACACCTACCTTAACC 27

```

RESULT 4

BM462502 1145 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT_6426189 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5518183
 DEFINITION 5' mRNA sequence.

ACCESSION BM462502
 VERSION 1
 KEYWORDS EST.

SOURCE BM462502.1 GI:18511542

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 1145)
 NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM12178 row: j column: 08

High quality sequence stop: 575.

Location/Qualifiers

1..1145

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5518183"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_71"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: NotI;

Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.1 kb.

299 a 318 c 273 g 255 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 2.94 Length: 1145
 Score: 112.50 Matches: 88
 Percent Similarity: 35.71% Conservative: 72
 Best Local Similarity: 19.64% Mismatches: 150
 Query Match: 3.43% Indels: 139
 DB: 12 Gaps: 21

US-09-856-327-2 (1-618) x BM462502 (1-1145)

QY 133 SerAlaProGlySerSerAlaHisLeuSerAsnGlyLysAsnPro-HisGlnArgGluPh 152

DB 28 AGTGGCCCGGTGGTCCACCGCGGTGGCGCGGTGGGACCTGCACAGCCCGGAG-- 85

QY 152 eGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerHisTrpTh 172

DB 86 -----AGACCCGTAGGC----- 97

QY 172 rCysSerThrProArgIleHisProMetGluSerLeuProGlyIleGlyArgProly 192

DB 98 ----AGGACCCCGGGA---CACCCCTGGAAGCAGGAAATGGACTCGGCTCTGAGGA 150

QY 192 sLeuSer-----AsnAspProAlaGluAspAsp 202

DB 151 TGTGGCTGTGAACCTTCACACTGAGGAGTGGGCTTTGTGGATCCTCCACG---AAGAA 207

QY 202 sGlnTrpAsnGluLeuTrpSerGluAlaGluArgLeuIleGlyThrSerThrLysGluPh 222

DB 208 ACTCTACAGAGATGTGATCGGGGAACCTTTAGAACCTGGCCTGTGAGAAAAAATG 267

QY 222 eAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspAr 242

DB 268 GGAAGAC-----CAGAGCATTTGAAGATTGGTACAAAATCA 303
 QY 242 gGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGly 262
 DB 304 GGGGAGAAATTTAAGAAAT-----CATATGGGAAGAGGACTCAGTGAAG 348
 QY 262 rValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLysGlnly 282
 DB 349 TAAAGAAATATGATCAATGTGGAGAACCTTCAGTCAGATTCTCAATCTTAATCTGAACA 408
 QY 282 sLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly----- 300
 DB 409 GAAAATTCCTACTATAGTAAGACCATGTGAATGT-----AGTTTGTGGGAAAGT 459
 QY 301 -----GlyTyrGluLys-LysIleGlyAlaAlaGluVal-----ArgAsnLeuL 315
 DB 460 CTTTCATGTCATCATTCATCCCTTAGTAGGCATCATGATCTCACCTGGGACACAAACCATTA 519
 QY 315 euAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLysValTrValL 335
 DB 520 TGACTATCAGGAAT-----ATGGAGAGAAACCATATATA 552
 QY 335 euAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuG 355
 DB 553 ATGTAAGC---AGTGTGGGAAAGCCTTCAGTTCTTGTCAATCCTTTTCGAAGACATGAAG 609
 QY 355 InValThr---ProArgAsnAspSerLeuIleProAsnLeuGly----- 368
 DB 610 AACTCACACTGGTGAGAAACCCCTATGCATGTCCGGAATGGGAAACCCCTTCATTTTC 669
 QY 369 -----ArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValL 383
 DB 670 TCTCCCAAGTGTCCCAAGACAACCTTGATTTAGCACCCCTGGGAGATGGACCTTATAA 729
 QY 383 euArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProTrpTrpLysG 403
 DB 730 TTGGTCAGGAATGG-----GGGGAACACCCCTTTGGAATGCC----- 767
 QY 403 luAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIleProPheArg 423
 DB 768 -----CCACAGTCAATCTCCGAATACCTTGACAAG 798
 QY 423 spProGluProGlnValThrThrProPheThr-----GluGluHisProt 438
 DB 799 AACCCCTCTCTGGGAAACAAACCCCTCTCCATGCGCGGTATGTGGCCCAAAACCTT 858
 QY 438 rpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspS 458
 DB 859 TTCACCTTTTACTC-----TGSCCAATAGCGGCCCTCCCTACGCGCTGC 873
 QY 458 erArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuL 478
 DB 874 CCCGATTTGTCCCAACACACAGG-----CCCTCCCGGACCT----- 911
 QY 478 euValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrA 498
 DB 912 -----TGSCCAATAGCGGCCCTCCCTACGCGCTGC 942
 QY 498 rgProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValA 518
 DB 943 TGTACTCTCCGGAACATCGCGGCACGCGCCGCCCAAAATTTGATTTGGCCGAGCAATT 1002
 QY 518 laSerAsnLeuGlyGlyTyrLeuProThrSerProGlnPheMetAspProGlyLeuA 538
 DB 1003 CCTCCACATCATGTTCTCATCTCCCGGGGAAT-CCA-----T 1040
 QY 538 laLeuHisLeuAlaGlyThr 544
 DB 1041 CACCACATCTGTCGGGTACC 1060

RESULT 5

BG541504

Fax: 301-838-0208
Email: cotown@etigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES

source	i. 839	Location/Qualifiers
/organism="Brassica oleracea"		
/mol_type="genomic DNA"		
/strain="T01000DH3"		
/db_xref="taxon:3712"		
/clone="BOGXY12"		
/clone_lib="BOGx"		

/notes=Vector: pHSO1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHSO1 using BstXI linkers"

BASE COUNT	250 a	136 c	221 g	232 t
ORIGIN				

Alignment Scores:

Pred. No.:	2.14	Length:	839
Score:	112.00	Matches:	59
Percent Similarity:	32.10%	Conservative:	28
Best Local Similarity:	21.77%	Mismatches:	78
Query Match:	3.41%	Indels:	106
DB:	28	Gaps:	12

US-09-856-327-2 (1-618) x BH451002 (1-839)

Qy	183	GlusSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluasp---Asp	201
Db	8	GAATTCCTCGTGGCGGAATAACCTCCATCGAAGCACAATAATTCAAGACCTGCAG	67
Qy	202	LysGluTrpAsnGluThrSerGluAlaGluArgLeuGlyThrSerThrylsGlu	221
Db	68	AGACATTGGAGAGGTCCAATCCGACACGATAG-----TCGGAAGAAGAG	115
Qy	222	PheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrtlysasp	241
Db	116	GTTTTTCAG-----GTTTCTAGAAATTGCAGGAAGCATATAAAGAT	157
Qy	242	ArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGlu	261
Db	158	GAGGAG-----GAG	166
Qy	262	TyrValGluTrpHis---SerAlaGluAsnLeuPheHisSerIleTyrtysAsnAspLys	280
Db	167	TAC-----TGCGATCAGAAAAGTCCAATATGTCG-----TATTCATCTCGGGAT	211
Qy	281	GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly	300
Db	212	CTAACAAATTTTATCATGCCCTTGACAAAGCAAAGCGGTTTCGGAAATAGGATTGGGT	271
Qy	301	GlyTyrGlu-----LysLysIleGlyAla	308
Db	272	CTTCATCATGAGCGGGTACTTCGATAACAGAGAGAGGTTGTGGAAAGGTGCGCTGTG	331
Qy	309	AlaGluValArgAsnLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrtle	328
Db	332	GAAATTTTTGAGGATTTATTTAGTACCACCTCTCCCA-----	367
Qy	329	MetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsn	348
Db	367	-----	367
Qy	349	SerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGly	368
Db	368	TCAGGTTT-----GATAGTTTCTTCGCGAGTAGGC	400
Qy	369	ArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnCluPheVal	388
Db	401	ATGCTATTACTCTCCFCAG-----ATGAATCAAGAGCTATTAAAGGCGAGCACAGAG	451
Qy	389	AspSerValArgAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGlnHis	408

MEDLINE PUBMED REFERENCE AUTHORS

21085660
11217851

5

TITLE

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II team.

JOURNAL

Nature 420, 563-573 (2002)

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

```
1. .1451
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4833441J24"
/db_xref="taxon:10090"
/clone="4833441J24"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
329. .1192
```

CDS

```
/note="unnamed protein product; HOTTL PROTEIN homolog [Homo sapiens] (SPTRIQ9UI99, evidence: FASTV, 85.8%ID, 93.4%length, match=821)
putative"
```

```
/codon_start=1
```

```
/protein_id="BAC26459.1"
```

```
/db_xref="gi:26325410"
```

```
/translation="MGLRNAKIHVERAVKOKIFMIGHYVPVIRCLLRQGWVEKMHVPPCTALPAQKDLSDSMLGNDSATEDDEENEFRESOLLDDGFLPFDLDLGIH ALMSRMVNETPYLLIWTTRDVLDCRFLSKDOMINHYARAGSFTTKVGLCLNRLPW FDEADADFFPCRYRLGAEDDKKAFPGKTPIPGLAPSCCLSHRGEKQASPCGLNPS PFLIPMGLPFTISQPMHSEKPREQGLYPVAGSRGGMVGSWLSLKLALPGSGTKNGSL RQSKFLFOHGF"
```

```
1438. .1443
```

```
/note="putative"
```

polya_site

```
1451
```

polya_site

```
/note="putative"
348 a 364 c 398 g 341 t
```

ORIGIN

Alignment Scores:

Pred. No.: 8.87 Length: 1451

Score: 109.00 Matches: 81

Percent Similarity: 34.92% Conservative: 44

```
Best Local Similarity: 22.63% Mismatches: 121
Query Match: 3.32% Indels: 112
DB: 11 Gaps: 22

US-09-856-327-2 (1-618) x AK029462 (1-1451)

QY 178 lleHisProMetGluSerLeuProGly-----IleGlyArgProLysLeu 193
:::|||||:::
Db 461 GTGCATCCCGCAGGCACGCCCTGCCAGCACCCAGAGGATCTCGATAGTCCATGCTG 520
194 SerAsnAspProAla-----GluAspAspLysGluTrpAsnGluLeuTyrSerGluAla 211
||| ||| :::::|||||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 GGTAAATAGCGATGCCAGCAGGATGAGGATGAGAGAGAGAGAGAAATGTTTGGGATCG 580
212 GluArgLeuIleGlyThrSerThrLysGluPheAspGlu---SerIleArgHisThrLeu 230
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 CAGCTGTTGGACTTGGATGGTTCCTGGAAATTTGATGACCTAGATGGGATACATGCTTTG 640
231 ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuPro 250
::: |||
Db 641 ATGTCCTCCG----- 649

251 LeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsn 270
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 -----ATGTTTCGAAATGAGACCCCTACCTCTACCTGACCACTCGCGGGAT 697
271 LeuPheHisSerIleTyrAsnAspLysGlnLysLeuPheThrLeuLeuThrAsn 290
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 698 GTGCTGGATTGCGTCTCTTCCAAGGATCAG-----ATGATAAAC 739
291 HisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGly---AlaAla 309
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 740 CAT-----TATGCCGTCGAGCTCTTCACCAAAAGGTGGCCTTTGCTCTC 787
310 GluValArgAsnLeuAlaThrArgAsnPro-----SerSerGlnLeuAspSer 326
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 788 AACCTCCGGAATCTG-----CGTGTGTTTGTAGAGCGCATGCTGACTCT 832
327 TyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeu 346
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 833 TTC---TTCCACGATGTACCGATG-----GGGGCAGAGGATGACAAAGAACCTTC 883
347 TyrAsnSerGlyPheSerGlyLeuGlnValTrpProArgAsnAspSerLeuIleProAsn 366
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 884 ATAGTAAGGG-----ACTCCA-----ATCCACGCG 910
367 LeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGlu 386
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 911 CTGGCT-----CCCTCATGCTGCTCTCCACACAG----- 940
387 PheValAspSerValArgAspProTyrGlyLeuProTyrTrpLysGluAlaValAla 406
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 941 -----GGAAAGACCAAGCTCCCTCTGTTGGCCTC----- 970
407 GlnHisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluPro 426
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971 -----TGTAAACCCCGCCCTTCTCTCTACCA----- 997
427 GlnValThrThrProPheThrGluGlnHisProThrHisThrGlnIleHisArgAspAla 446
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 998 ---ATGGGTCTTCCATTCATAGGCCAA---CCTTGGCATTCAGAAAGCCTAGGAGCAA 1051
447 PheSerTyr-----GlyAlaValGlyProGluValAspSerArgValIleValAsp 463
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1052 GCCTGTACCTGTAGCGGGTCTAGAGGTGGGTGGTGGTCTTCTGCTGGAGTCTAGAT 1111
464 LeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuLeu----- 478
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 AAGTGTCTGGTGGACCGAGGTCTACAAAGCAATGGTGTCTTAAGCAAGTAAGTTC 1171
479 ValPheGlnAsn-----AspValGlnAspGlyTyr--- 488
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1172 CTGTTTCAACATGGGTTCTGAGTTTTCAGAAATTTGGATTCAGATCAGACGCTTTTAA 1231
```



```

Db 2723 -----AGACCCATGAGCGGTGAGCATCTGACGCTGGTCCCATGCTCTTTCCAGG 2775
Qy 419 ePro-----PheArgAspProGluProGlnValThrThrProPheThrGluGluHis-- 436
Db 2776 GCCAGGACTATTCTAGTGTAGTCCCTTACATGGAAACAGCCACGACTGGAGGAACCATGA 2835
Qy 437 ----ProTrp 438
Db 2836 GGGGCGCTGG 2845

RESULT 9
BJ258831
LOCUS
DEFINITION BJ258831 690 bp mRNA linear EST 08-APR-2002
aestivum cdna clone whh9m05 5', mRNA sequence.
ACCESSION BJ258831
VERSION BJ258831.1 GI:20081455
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 690)
Expressed genes in Triticum aestivum
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..690
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whh9m05"
/tissue_type="spike at heading date"
/dev_stage="Peekes", scale 10.5"
/clone_lib="Y. Ogiwara unpublished cdna library, Wh_h"
BASE COUNT 159 a 190 c 189 g 152 t
ORIGIN

Alignment Scores:
Pred. No.: 4.99 Length: 690
Score: 107.00 Matches: 51
Percent Similarity: 40.94% Conservative: 19
Best Local Similarity: 29.82% Mismatches: 64
Query Match: 3.26% Indels: 37
DB: 12 Gaps: 9

US-09-856-327-2 (1-618) x BJ258831 (1-690)

Qy 120 ArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProGlySerSer 139
Db 77 AGGAGGAGTGGCGTCCG-----TCCCTCCCGGAGCAGTCC 115
Qy 140 AlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAla 159
Db 116 GCATCTCCGCTCAAGGCGCAGGAGTTACGGGACAGCGGAGTACAAAGTCTTACG 175
Qy 160 ValThrArgGlyValGlyMetSerThrHisTrpThrCysSerThrProArgIleHis 179
Db 176 GTGACG-----GGCGGTACAACTCGCTGGTGGTGGGAAACCCCTGTACAGT 223
Qy 180 ProMetGluSerLeuProGlyIle---GlyArgPro-----LysLeuSerAsnAsp 196
Db 224 CCACCAACCAAAATTCGACTCGGCTGGCTGGCGGCTTCTTCGAGGGCTCCCTGGAG 283

```

```

Qy 197 Pro-AlaGluAspAsp-----LysGluTrpAsnGluLeuTyrSerGluAlaG1 212
Db 284 CCATACACAGGACGCCCTGACCCCTGATGAGGAGGTAGAGTACATGCATGCCATCTTGGC 343
Qy 212 uArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuVal 232
Db 344 CGGGCATTTGGGACATGTGTTCAAGG----- 371
Qy 232 uArgSerLeuGlnAspAlaTyrLys-----AspArgGlnArgIlePheArgProLe 249
Db 372 -AGAAGCTTCAACAGCGCTACCGATGAGCCGCTACGCTGGTCAACAGCGTTTCCATCAAGT 430
Qy 249 uProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAlaG1 269
Db 431 TCAC---TCCTGCTCTCTCTGTAATCCGGCCAGTTCTCGAGTCATCTGG-----AT 478
Qy 269 uAsnLeuPheHisSerIleTyrAsnAspAsp 279
Db 479 GAACCTGTTTAAACAATGCAATAATGAT 509

RESULT 10
BU518146
LOCUS
DEFINITION BU518146 1319 bp mRNA linear EST 12-SEP-2002
AGENCOURT_10166431 NIH_MGC_134 Mus musculus cdna clone
IMAGE:6515841 5', mRNA sequence.
ACCESSION BU518146
VERSION BU518146.1 GI:22825672
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1319)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14093 Row: g Column: 10
High quality sequence start: 177
High quality sequence stop: 557.
Location/Qualifiers
1..1319
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6515841"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_134"
/notes="vector: pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2:
NotI; Cloned unidirectionally. Primer: Oligo dr. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."
BASE COUNT 277 a 400 c 331 g 302 t 9 others
ORIGIN

Alignment Scores:
Pred. No.: 12.1 Length: 1319
Score: 107.00 Matches: 105
Percent Similarity: 31.60% Conservative: 47
Best Local Similarity: 21.83% Mismatches: 174
Query Match: 3.26% Indels: 158
DB: 13 Gaps: 25

US-09-856-327-2 (1-618) x BU518146 (1-1319)

```


Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seyer, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weltz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.

TITLE
Functional annotation of a full-length mouse cDNA collection

JOURNAL
Nature 409 (6821), 685-690 (2001)

MEDLINE
21085660

PUBMED
11217851

REFERENCE
5

AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE
Analysis of the mouse transcriptome based on functional annotation

JOURNAL
Of 60,770 full-length cDNAs

REFERENCE
Nature 420, 563-573 (2002)

AUTHORS
6 (bases 1 to 3111)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE
Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

FEATURES
Location/Qualifiers

1..3111

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:D730004B05"

/db_xref="taxon:10090"

/clone="D730004B05"

/sex="female"

/tissue_type="mammary gland"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="10 days lactation, adult"

311..2807

/note="putative"

misc_feature
tenascin C (MGD|MGI:101922, GB|NM_011607, evidence:

BLASTN 98%, match=817"

3093..3098

/note="putative"

polyA_site
3111

BASE COUNT	733 a	794 c	910 g	674 t
ORIGIN				
Alignment Scores:	44.1			Length: 3111
Pred. No.:	106.50			Matches: 94
Score:	35.00%			Conservative: 53
Best Similarity:	22.38%			Mismatches: 148
Query Match:	3.24%			Indels: 125
DB:	11			Gaps: 24
US-09-856-327-2 (1-618) x AK085667 (1-3111)				
Qy	246	PheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrp	265	
Db	2810	TTCTCACTGATGCAGATGCTGTGTGACAGGGTATGCTATAAGCCAGAGTTCGCAAGAGA	2751	
Qy	266	HisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPhe	285	
Db	2750	CACAAT-----CAGCTCTATCACAGGCTTTGG-----	2724	
Qy	286	ThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly---GlyTyrGluLys	304	
Db	2723	-----CTGGTACTCAGGGTTGTGTACTTCTTCTT	2694	
Qy	305	LysIleGlyAlaAlaGluVal-----ArgAsn-----LeuLeuAlaThr	317	
Db	2693	CAACCGGGGCTCGGGTGTGTTTTCACAAATGTGCAGAGATATTTCTACTCTTTGGCC	2634	
Qy	318	ArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSer	337	
Db	2633	AGGACGAAGGCCAGT-----TTGGCGGTA	2610	
Qy	338	GlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThr	357	
Db	2609	GGAGTGCTCTGCTCTCTCAGACTTTTGTGTCTCTCCCTCATCTTCTTGTTCATATT	2550	
Qy	358	ProArgAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAla	377	
Db	2549	TCGGAATGAT-----CTCCCATGTTTCAAGCAATGTCTAGAGATGCCACTCTAC	2496	
Qy	378	PheCysGlnIleValLeu---ArgGlnGluPheValAspSerValArgAspProTyr	396	
Db	2495	TTCCACAGATCTCTCTTGTATAGACTTCAATTT---TAGACCTTCAGG-----	2451	
Qy	397	GlyLeuProTyrTrpLysGluAlaValAlaGlnHisIleAlaLysAsnPro---ThrAsp	415	
Db	2450	-----TGCAGGCAAAATAGGTGGCAACCTCGGCACTGAC	2418	
Qy	416	Ala---LeuProIleProPheArgAspProGluProGlnValThrThrProPheThrGlu	434	
Db	2417	AGGGATGCTCTCTTGTGTTTCCAGATGGCGAA-----CACACGAAT	2376	
Qy	435	Glu-----HisProTrp-----His	439	
Db	2375	GAAGTACTCCACCCTGGTTCCAGTCCCGAATGGTGGTGGTCTGGTCCCGGCGAC	2316	
Qy	440	ThrGlnIleHisArgAspAlaPheSerTyr---GlyAlaValGlyProGluVal-----	456	
Db	2315	ACGGAATGTCATCTCTAGGCCATCAGCATGGGTGGGTGTGTACATAATGAGTACTCAGT	2256	
Qy	457	AspSerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsn	476	
Db	2255	GACCCGGCATCTCATTTGTCATGCGAGATTACAGTCTCCCTCTGTTACTTCTGTCACAAT	2196	
Qy	477	LeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArg	496	
Db	2195	AAGTCTTTGGAGGGGACACCTCAGACGACTCTAT-----CCCTGTGTACAC	2148	
Qy	497	TyrArgProSerThrAlaSer-----	503	
Db	2147	TTTATTGCAGATCGAGGGGCTGACACACATTTGCTTGTGTTGCTGCTGATGGGACA	2088	

QY 504 -----AsnValArgAlaArgLys-----MetMetAlaAspMetCysGluValAla 518
|||:::|||||||:::|||||||
Db 2087 GGAGCGCTGCCACAGTCCAGCCGTAAGCCCTCATGGCAGATACACTGGCGGCTCTC 2028
QY 519 SerAsnLeuGlyGlyTyrLeuProThrSerProPro-----530
Db 2027 ACAGCGCGCTTGCCCATGCAATCACTGGGGCACCTTTGCTCTTTCGACGTCTTGCCAGT 1968
QY 531 ---GlnPheMetAspProGlyLeuAlaLeuHisLeuAla-----GlyThrThrArgIle 547
Db 1967 GAAGCCCTCGTGGCAGATCGATGGCCATTCACAGCGGCCGATGCCCTGGCAGTCACT 1908
QY 548 GlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTirPaspPheAlaAsn 567
Db 1907 GGGCGAGGAAGCTCAGCAGACAGTCAAGGCCATCTCACACATGCAGTGTCC 1848
QY 568 LeuTyrValAla-----GlyAsnGlyThrIleArgThrGlyPhe 580
Db 1847 GTCTACACAGCGCCCGCTGGCTACAGTCCCGGGGACAGCGCGTCTCTCGACGTCTTC 1788
QY 581 GlyGluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIle 600
Db 1787 CCAGTGTAGTCTCATCAGAT---ACACATGCCATTTCACACAGCGGCCATGTTGGTG 1731

RESULT 12

AV764246 691 bp mRNA linear EST 19-OCT-2000
LOCUS AV764246 MDS Homo sapiens cDNA clone MDSGB05 5', mRNA sequence.

ACCESSION AV764246

VERSION AV764246.1 GI:10922094

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 691)

Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H.,

Gu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng

L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G.,

Yang, Y., Gao, G., Zhang, Q., Chen, S., Han, Z. and Chen, Z.

Homo sapiens cDNA MDS clones

Unpublished

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

Location/Qualifiers

1..691

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MDSGB05"

/tissue_type="Bone marrow"

/cell_type="CD34+ hematopoietic stem/progenitor cell"

/lab_host="PM25.8"

/clone_lib="MDS"

/note="Vector: pTriplex2; Site_1: sfliA; Site_2: sfliB"

BASE COUNT 178 a 176 c 173 g 164 t

ORIGIN

Alignment Scores:

Pred. No.: 6.25 Length: 691

Score: 106.00 Matches: 69

Percent Similarity: 37.97% Conservative: 51

Best Local Similarity: 21.84% Mismatches: 94

Query Match: 3.23% Indels: 103

DB: 9 Gaps: 15

US-09-856-327-2 (1-618) x AV764246 (1-691)

QY 152 PheGluAsnLeuSerAla-----GluAlaValThrArgGlyValGlyGlyMetSerThr 169
|||:::|||||||:::|||||||
Db 15 TTTAATCACTCTGCAGTCTTGGAGAGCAGGTTCAAGAGCTGCAGCGGGGCCACTATA 74
QY 170 HisTrpThrCysSerThrProArgIleHisProMetGluSerLeuProGlyIleGly 189
:::|||||||
Db 75 TATAGATGTTGTCTC-----AGGAGCTCCCGCAGGAATGGA 110
QY 190 ArgProLysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrSer 209
:::|||||||
Db 111 GTTCACAGAAAGAGT---GACCT---GAGGAACCTAAAGGAACCCAGCAAGCTTTCT 164
QY 210 GluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThr 229
|||||||
Db 165 GAAGCTCAGCAGCAGCTATGCAACACAGAGGAAGTGAATGAA-----209
QY 230 LeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePhe-----Arg 247
|||||||
Db 210 -----TTAAGGAAGCTGCTGGAAGAAGAACAGAC---CAAAGAGTGGCTGCTGAGAAT 260
QY 248 ProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSer 267
|||:::|||||||
Db 261 GCTCTCTGTGGCGAGAGCAGATCAGACGGTTA---GAGCAGAGTGAATGGGACTCT 317
QY 268 AlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeu 287
:::|||||||
Db 318 TCCCGGACTCTCTATCATTTGGCTCCTGTGGCACTCAGGAGCGGCACTGTTATAGAT--- 374
QY 288 LeuThrAsnHisArgCysThrArgLeuAlaLeuThrGlyGlyGlyGlyGlyLysIleGly 307
|||||||
Db 375 CTTACAAGCAACAGTGTGCGAAGGACCCGAGTGGCTGGATGGAAGCGAGTCTCTG--- 431
QY 308 AlaAlaGluValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyr 327
|||||||
Db 432 -----CGTTCATCTTCTGCACTTCCAGGAC---CCGAGTGCCCACTTCTAGCAGCCAT 478
QY 328 IleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyr 347
:::|||||||
Db 479 CTACTTCT-----AATGATTTCATGCTCTGCT 505
QY 348 AsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeu 367
:::|||||||
Db 506 CATCTCTGTGTTTACGGC-----523
QY 368 GlyArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPhe 387
:::|||||||
Db 523 -----523
QY 388 ValAspSerValArgAspProTyrGlyLeuProTrpTrpLysGluAlaValAlaGln 407
Db 523 -----523
QY 408 HisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGln 427
|||||||
Db 524 CATCTATAGACTTAGTGTCTACTCTTGCACCACCTCCCTTA-----CAACTTGAA 574
QY 428 ValThrThrProPheThrGluGluHis-----ProTrpHisThrGlnIleHisArgAsp 445
:::|||||||
Db 575 ATTCTCTCACCCTTAACATCAGAACATCAATTTCTAGTGGGAACAGCGTCTCCATTACAG 634
QY 446 AlaPheSer-----TyrGlyAlaValGlyProGluValAspSer 458
|||||||
Db 635 GCCTTCTCTCCAACCTTTTCCCGGAGCGCTGCACCCGAGGTGATACC 682

RESULT 13

BH494670/c

LOCUS

DEFINITION

BOJ JV59TR BOJ Brassica oleracea genomic clone BOJ JV59, genomic

survey sequence.

BH494670

ACCESSION

BH494670.1 GI:17702774

Alignment Scores:

Pred. No.:	9.26	Length:	920
Score:	106.00	Matches:	79
Percent Similarity:	34.78%	Conservative:	41
Best Local Similarity:	22.90%	Mismatches:	128
Query Match:	3.23%	Indels:	98
DB:	10	Gaps:	12

US-09-856-327-2 (1-618) x BE782505 (1-920)

QY	123	AsnValProThrLeuAspProGlyAlaTrpSerAlaProProGlySerSerAlaIleSer	142
DB	846	CATATCCAGACTATCAAGAGGTCTAAGCCAGCTCCACCTTTTGCTGCTTGGTAA	787
QY	143	AsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSer	156
DB	786	AATGTAGACTTCCAAATTTGAGCCGAGCTGTATTCATCTTTGGGCCACATTTCTCT	727
QY	157	---AlaGluAlaValThrArgGlyValGlyGly-MetSerThrHisTrpThrCysSerTh	175
DB	726	GGGTCCAAAGGCTCTCGCACTCTACTTCTGCTCTCAGCGTTTCTGCCACTTGTCCAC	667
QY	175	rProArgIleHisProPrometGluSerLeuProGlyIleGlyArgProLysLeuSerAs	195
DB	666	TCCTCTTGTTT-----ATTCAACTCCGCAATATATCTCA	631
QY	195	nAspProAla-GluAspAspLysGluTrpAsnGluLeuTySerGluAlaGluArgLeuI	215
DB	630	GAAGAGTGCCTGAA-----TGTCGTCTCATTCACGAGTAG	595
QY	215	leGlyThrSerThr-----LysGluPhe-AspGluSerIleArgHisThrLeuValLeu	232
DB	594	TTTCAACAAGAGTCTCTCTCATTTTCAGATTCTCGGTACCACATCTTTGGGGCTA	535
QY	233	ArgSerLeuGlnAspAlaTyrlsAspArgGlnArgIlePheArgProLeuProLeuAla	252
DB	534	AG-----GTATACCCGCAAGCCCAATGATGCTCGCATTAATGTTGC	487
QY	253	CysHisArgLeuLysAsn-AlaProGluTrpValGluTrpHisSerAlaGluAsnLeuPh	272
DB	486	TGTCACGCGCGGTAAATCATACAGAAATATGTTCTGTGCACAGGCTTCGGGCCAAC	427
QY	272	ehisSerIleTyrlsAsnAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisAr	292
DB	426	AAATCCCGTTTAC-----AGGCAGGAGCTTCACAGAAATCGG-ACAAACACAC	382
QY	292	gCysThrArgLeuAlaLeuThrGlyGlyTyrlsLysLysIleGlyAlaAlaGluValAr	312
DB	381	TGTACTCGAGCAGCTTCTTAGGAGGAGGATCAGAG-----	346
QY	312	gAsnLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrlleMetAlaLysVa	332
DB	345	-----TCATCTTCGACAACTCAAGCAATTTAGCACATAC-----	310
QY	332	lTyrrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrlsAsnSerGlyPheSe	352
DB	309	-----GGGTTTTC	302
QY	352	rglyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrlleTh	372
DB	301	A-----TCATCATCCCTTCAATCAACACAACTGTCCAATCTGGATGTGAATCTCGGT	248
QY	372	rgluGlnProMetAlaPheCysGlnIleValLeuArgGln-----GluPheValAs	389
DB	247	GGAACAACCTTCTGTTTTCACACACATTTCTATAGTTTGGTAGTGCAGTTTTCGATC	188
QY	389	pSerValArgAspProTyrlsGlyLeuProTrpTrpLysGluAlaVal-----	405
DB	187	CACAAGGCGCTGCCAACCCATGAATAAGTTTCTGCTTTCAGCCTTGTGGGTAGTGT	128
QY	405	-----AlaGlnHi	408
DB	127	GCCATGCCGGCTCCAAAGCTTCTCCGGCTAGCCGCCAAGCCCTCCGTTTCTTACCGCCTC	68

QY 408 sileAlaLysAsnProThrAspAlaLeuProfileProPheArgAspProGluProGlnVa 428
Db : : : : : ||| ||||| ||| : : : : :
67 CACTTCGAGAGTCGCGGGGATCGCGACCTCTAACCTGCACACAGCGCTCTCTCTAC 8
QY 428 lThrThr 430
Db : : : : : ||| ||||| ||| : : : : :
7 AACTACG 1

RESULT 15
BC034946 4866 bp mRNA linear HTC 04-MAR-2003
LOCUS Homo sapiens, Similar to secretory pathway component Sec31B-1,
DEFINITION clone IMAGE:4826447, mRNA.
ACCESSION BC034946
VERSION BC034946.1 GI:23273821
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4866)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Paikovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 33 Row: 0 Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14591929
This clone has the following problem: retained intron.

FEATURES
source

1..4866
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4826447"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
BASE COUNT 1104 a 1289 c 1294 g 1179 t
ORIGIN

Alignment Scores:

Pred. No.: 91 Length: 4866
Score: 106.00 Matches: 81
Percent Similarity: 34.05% Conservative: 46
Best Local Similarity: 21.72% Mismatches: 137
Query Match: 3.23% Indels: 109
DB: 11 Gaps: 18

US-09-856-327-2 (1-618) x BC034946 (1-4866)

QY	229	ThrLeuValLeuArgSerLeuGlnAspAlaTyrlsAspArgGlnArgIlePheArgPro	248
DB	2459	ACAGATCTGCTGAAGCAACACACAGAGCGCTACTTGGCCCAAGAGAAACCAAAATCTCC	2518

Search completed: September 24, 2003, 08:50:25
Job time : 3544 secs

```
Qy 249 LeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrVal----- 263
Db      |||||
Db 2519 TCGCTTCTAGCTGTGTGTCACAAAGAAATGGAAGGATGGTGTGTACCTGTAGCCTG 2578
Qy 264 ---GluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLys 282
Db      |||||
Db 2579 AAGAACTGGAGAGGAGCTGGCTTGTCTACTGACATACTCAGGCACAGAGAAATTTCCC 2638
Qy 283 LysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeu----- 296
Db      |||||
Db 2639 GAGCTCTGTGACATGCTGGGA-----ACTGCGATGGAAACAGGAGGGCAGCAGG 2686
Qy 297 AlaLeuThr-----GlyGlyTyrGluLysLysIle 306
Db      |||||
Db 2687 GCACATAACCTCCGAAGCCAGACTCTGTATGTGTGTCAGGGAGTGTGGAGCGGTGCTG 2746
Qy 307 GlyAla---AlaGluValArgAsnLeuLeuAlaThrArgAsnProSerSerGlnLeuAsp 325
Db      |||||
Db 2747 GAGTGTGGGGCAAAATGCCACGAGCTTTG-----TCCCCATGGCTCTGCAGGAC 2797
Qy 326 SerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIle 345
Db      |||||
Db 2798 -----CTGATGGAGAGGTGATGTTCTT-----AACAGGAGCTTG 2833
Qy 346 LeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIlePro 365
Db      |||||
Db 2834 GAGCAACTGCGGGGTCTCATGTGGGTGAGCCAGCCCTGCCACAACCTACAGGGTCACT 2893
Qy 366 AsnLeuGlyArgTyrIleThrGluGln-----PrometAlaPheCysGln 380
Db      |||||
Db 2894 CAGTATGCCAACCTCCCTGGCAGCCAGCCAGCCAGCTGGCCACTGCCCATGAGCTTCTACCC 2953
Qy 381 IleValLeuArgGlnGluPheValAspSerValArgAsp----- 393
Db      |||||
Db 2954 AGGCACTGTGCTAGCCACCAGTTCAGCAGTAAGAGATCGGCTTTTCATGCTCAAGGT 3013
Qy 394 -----AspProTyrGlyLeuPro----- 399
Db      |||||
Db 3014 TCTGTGTGCTTGGGCCAACAGTCTCCCGCTTCCCTTCCCGCGATTGTTGTGGAGCT 3073
Qy 400 -----TrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
Db      |||||
Db 3074 ACCCTCCACTTAAGAGACATCATCTACAGATTGGGATCCCGCCCTTCTCACCAGGTC 3133
Qy 418 ProIleProPheArgAspProGluProGlnValThrProPheThr-GluGluHis-- 436
Db      |||||
Db 3134 CCAACTCCA-----TCTCCAAGCCCAAGGGTTTTTACCCCTCAGTCACTCACCAGCGATG 3187
Qy 437 -ProTrpHisThrGlnIle-----HisArgAspAlaPheSe 448
Db      |||||
Db 3188 CCCTTGGCACCTTCCCATCTCCTAGCCCTTATCAGGGTCCCGAGGACACAGATATAAGTGAC 3247
Qy 448 rTyrGlyAlaValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheG 468
Db      |||||
Db 3248 TACAGGGCACCTGGGGCCCCAG----- 3269
Qy 468 yAlaThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsnAspValGlnAspGlyTy 488
Db      |||||
Db 3270 -GCCATCCAGCT-----TTGCCTTTGAGCCCTGGGGTAAGGCTGGA-- 3311
Qy 488 rSerMetProGlnProPheArgTyrArgProSerThrAlaSerAsnValArgAlaAr 508
Db      |||||
Db 3312 -----CCTCAAGATCTCGGAAGAAGCCCGCCAGCCCGGGAACCTCCAGAGGAA 3364
Qy 508 gLysMet-----MetAlaAspMetCysGluValAl 518
Db      |||||
Db 3365 CAAGCTGCAGAGACATTTATGCCCCAGCACCAATTACTCTCCAGTTATGAGCCTCAC 3424
Qy 518 aserAsnLeuGlyTyrLeuProThrSerProPro 530
Db      |||||
Db 3425 CCCTGAGCTACAGGGATTTCTTCCTCAGACGCCCT 3461
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 04:17:00 ; Search time 461 Seconds
(without alignments)
3618.769 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MS1STEQMLRDYPRSMQING.....IINTLKGKGTGDKNTGHEHNL 618

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USP70_spool/US09856327/runat_23092003_153418_15498/app_query.fasta.1.775
-DB=N_Geneseq_19jun03 -QWRT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosmb2 -TRANS=human40.cdi
-LIST=45 -DICALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856327 -CGCN_1_1_312 -runat_23092003_153418_15498 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19jun03.*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3284	100.0	2106	22	AAF99980 Nucleotide sequenc
2	1232	37.5	1902	21	AAZ46411 Pleurotus cornucop
3	1222.5	37.2	1869	17	AAT34420 Pyranose oxidase e
4	1210	36.8	1869	22	AAC87519 Trametes hirsuta p
5	1210	36.8	1995	22	AAC87518 Trametes hirsuta p
6	1091	33.2	1701	20	AAV83626 Nucleic acid encod
7	1044	31.8	1946	21	AAA71487 T. matsutake pyran
8	1044	31.8	1946	21	AAA07403 Trichoderma derive
9	208.5	6.3	2467	24	ABK52367 DNA encoding gluco
10	207	6.3	4290	23	ABL52917 2-keto-D-gluconate
11	206	6.3	5187	20	AAK57912 G. oxydans D-sorbi
12	200	6.1	1632	20	AAK57909 G. oxydans D-sorbi
13	198	6.0	4665	20	AAK52345 Membrane-bound glu
14	174	5.3	1689	24	ABL52744 Glucose-3-dehydrog
15	159.5	4.9	2484	21	AAC36995 Arabidopsis thalia
16	148.5	4.5	1641	20	AAZ10823 Choline oxidase (C
17	148	4.5	4403765	22	AAI99683 Mycobacterium tub
18	148	4.5	4411529	22	AAI99682 Mycobacterium tub
19	124	3.8	2400	18	AAAT75000 Choline oxidase ge
20	124	3.8	2576	21	AAZ95701 Humicola insolens
21	124	3.8	2576	22	AAH47743 H. insolens DSM 18
22	122.5	3.7	1503	22	AAK32484 A. tumefaciens 1,5
23	122.5	3.7	1841	22	AAK32462 Choline oxidase ge
24	121.5	3.7	2400	17	AAAT42859 Candida tropicalis
25	121	3.7	4233	20	AAK87946 Listeria monocytog
26	121	3.7	2944528	24	ABA03041 Nucleotide sequenc
27	120.5	3.7	8918	21	AAZ36925 Stealth virus nucl
28	120.5	3.7	9199	20	AAK84322 Arabidopsis thalia
29	119	3.6	2336	21	AAC44869 DNA encoding novel
30	119	3.6	8526	23	AAK79203 Candida claocae fa
31	116.5	3.5	4305	20	AAK87945 erya region of S.
32	115.5	3.5	29879	14	AAQ46806 P. putida KT2440-a
33	115	3.5	1596	22	AAK60988 Mycobacterium tub
34	114	3.5	4403765	22	AAI99683 Mycobacterium tub
35	114	3.5	4411529	22	AAI99682 Mycobacterium tub
36	113.5	3.5	35829	23	AAS59573 Propionibacterium
37	113	3.4	8253	24	ABO71068 Listeria monocytog
38	113	3.4	21185	21	AAAG3350 Streptomyces globi
39	113	3.4	63164	21	AAAG3348 Streptomyces globi
40	112.5	3.4	2385	13	AAO27956 Cholesterol oxidas
41	112	3.4	1911	25	ABK56231 CDNA encoding corn
42	111	3.4	29379	23	AAS59510 Propionibacterium
43	111	3.4	53178	23	AAS59543 Propionibacterium
44	110.5	3.4	34980	24	ABO81848 Bifidobacterium lo
45	109.5	3.3	1818	21	AAZ48464 Glucose oxidase (G

ALIGNMENTS

RESULT 1

AAF99980

ID AAF99980 standard; CDNA; 2106 BP.

AC AAF99980;

XX 20-JUL-2001 (first entry)

DT 20-JUL-2001 (first entry)

XX Nucleotide sequence encoding Lyophyllum shimeiji antibacterial protein.

XX Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;

KW Pyricularia orizae; Rhizoctonia solani; rice pathogen; ss.

XX Lyophyllum shimeiji.

OS Lyophyllum shimeiji.

XX Key

FT Location/Qualifiers

CD 8..1864

FT /*tag= a
 XX /product= "antibacterial protein"

PN W0200121657-A1.

XX 29-MAR-2001.

PD 20-SEP-2000; 2000WO-JP06404.

XX 21-SEP-1999; 95JP-0267238.

XX (NIBB) JAPAN TOBACCO INC.

PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX Takakura Y, Kuwata S, Inoue Y;

XX WPI: 2001-281598/29.

DR P-PSDB; AAB97035.

XX Antibacterial protein and encoded gene isolated from *Lyophyllum*
 PT *shimeji*, with activity against plant pathogenic bacteria, applicable in
 PT agriculture e.g. rice cultivation at low concentration, produced at low
 PT cost on large scale .

XX Claim 12; Page 38-42; 52pp; Japanese.

CC The present sequence encodes an antibacterial protein from the fungus
 CC *Lyophyllum shimeji*. The protein was obtained from a fraction prepared
 CC by extracting *Lyophyllum shimeji* with water and subjecting the extract
 CC to ammonium sulphate precipitation. The protein inhibits the growth
 CC of the plant pathogenic bacteria *Pycularia oryzae* and *Rhizoctonia*
 CC *solani* at a relatively low concentration. *P. oryzae* and *R. solani* are
 CC causative of the two major diseases of rice. The protein contains
 CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
 CC The antibacterial protein can be produced at low cost on a large scale.

XX SQ Sequence 2106 BP; 593 A; 540 C; 527 G; 446 T; 0 other;

Alignment Scores:

Pred. No.: 2,066-279 Length: 2106
 Score: 3284.00 Matches: 618
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-856-327-2 (1-618) x AAF99980 (1-2106)

QY 1 MetSerLeuSerThrGluGlnMetLeuArgAspTyrProArgSerMetGlnIleAsnGly 20
 DB 8 ATGCTCTCTCAACCGAGCAGATGCTACGCGACTATCCACGGTCTATGCAAAATCAACGGA 67
 QY 21 GlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspValPheIle 40
 DB 68 CAGATTCTCTAAGACGCAATTCACGAAACATACGGAACACGACGGAGTTGATGTATTCATT 127
 QY 41 AlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAlaGlyLeuArg 60
 DB 128 GCAGGATCTGCNCCCATTTGGGCGACGATATCAAGCTCTGTGTGAAGCTGGTCTACGT 187
 QY 61 ValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGly 80
 DB 188 GTTGTGATGGTCGAGATCGGAGCTGCTGATAGCTTCTACGCTGTTAATGCGGAAGAGGA 247
 QY 81 ThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGluPheGlnLysAsp 100
 DB 248 ACTGCAGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 307
 QY 101 IleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSerValProValArg 120
 DB 308 ATTGACCGCTTCGTCATATGTAATCAAGGAGCCCTACAAACAGTCTCTGTTCTCTCAGA 367
 QY 121 AsnGlnAsnValProThrLeuAspProGlyAlaIlePheSerAlaProProGlySerSerAla 140

DB 368 AACCAAGACGTCCTTACACTTGATCCCGAGCCTGGAGCGGCCCTCGAAGTTCAGCC 427
 QY 141 IleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
 DB 428 ATATCGAACGGTAAATAATCCCTACCCAGCGGAATTCGAGAACTTGTCTGCGGAGGCCGTA 487
 QY 161 ThrArgGlyValGlyMetSerThrHisTrpThrCysSerThrProArgIleHisPro 180
 DB 488 ACGCTGGAGTCGGGGCATGATACCCACTGGAGTGTCTCCACGCCAGGATTCATCCA 547
 QY 181 ProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAsp 200
 DB 548 CCCATGGAAGTCTCCCGGCATCGGCCGCTCCGAAGCTCAGTACGACGCCGACAGGAC 607
 QY 201 AspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLys 220
 DB 608 GACAAAGAGTGAACGAGCTTTATTCGAGGCGGAGCTCTCATCGGACTTCCACCAAG 667
 QY 221 GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys 240
 DB 668 GAATTCGACGAGTCAATTCGGCACACACCTTGTCTGCGCTCTTTCGAAGACGGTACAA 727
 QY 241 AspArgGlnArgIlePheArgProLeuAlaCysHisArgLeuLysAsnAlaPro 260
 DB 728 GATCGTCAACGTATCTTTCCGCCCTCTCCCGTTGGCATGCCACCGGTTGAAGAACGCGCG 787
 QY 261 GluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLys 280
 DB 788 GAATACGTCGATGGCATGCACTACAGAAATCTTTTCCACTCTATCTACAGCATGACAA 847
 QY 281 GlnLysLysLeuPheThrLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly 300
 DB 848 CAGAAGAGCTCTTTACCTCTGCTGACGAACCATCGCTGCACACGACTGCGCTTACGGGC 907
 QY 301 GlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 320
 DB 908 GGGTATGAGAAAGATTGGCGCTGCGGAGGTCAGGAATCTACTGGCCACCCAGGAATCCT 967
 QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
 DB 968 AGTTGCGCAGCTGGACAGCTATATATGCGGAAGGTATATGACTGGCGCTCGGGACGATC 1027
 QY 341 GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn 360
 DB 1028 GGCAACCCACAGATTCTCTAATCTCGGCTTCTCTGGGCTACAGGTACGCCACGCAAT 1087
 QY 361 AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln 380
 DB 1088 GACTCGTTGATCCCAACCTGGGAGGTACATCATCGGAGCAGCCGATGGCATTTTGCAG 1147
 QY 381 IleValLeuArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProTrp 400
 DB 1148 ATAGCTTGGAGCGAGGAATTCGTGACAGCTGGCGGACGATCTTATGGAGTGGCATGG 1207
 QY 401 TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro 420
 DB 1208 TGGAAAGAACCGTGTGCTCAACATATGCCAAGAACCAGACAGATGCATGCCCATTCG 1267
 QY 421 PheArgAspProGluProGlnValThrThrProPheThrGluGluHisProTrpHisThr 440
 DB 1268 TTCGCGCATCGGAACCCAGGTAAACCCCATTTACAGAAGAACACCCCTGGCACACG 1327
 QY 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgVal 460
 DB 1328 CAGATTACCCCGATGCTTTTTCGTACGCTGCCGCTGCTGCTGAGGTGACATCTCTGCTGC 1387
 QY 461 IleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPhe 480
 DB 1388 ATCTGCGACCTCGCTGGTTTGGCGCAACCCAGCTGAAGCAACAACTTTTGGTTTC 1447
 QY 481 GlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSer 500

Db	1448	CAGAACGATGTTCAAGACGGGTACAGTATGCCCGACCGGACGTTTCAGATATCGACCCAGC	1507
Qy	501	ThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaLaserAsn	520
Db	1508	ACTCGGTCAAAACGTGAGAGCAAGGAAAAATGATGCCGATATGTGCGAAGTGGCGAGCAAC	1567
Qy	521	LeuGlyGlyTyrLeuProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHis	540
Db	1568	TTGGGAGGTATTGTCACAGTCCCGCCGACGTTTATGGATCATCAGGCGCTTGCACTTCAT	1627
Qy	541	LeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAlaAspAsnSer	560
Db	1628	CTTGGCGGGGACTACTCGCATGGCTTCGACAGGCAACTACAGTGGCTGATAACAACCTCG	1687
Qy	561	LeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPhe	580
Db	1688	CTGGTCTGGGACTTTGCCAATCTTTATGTTGCAGGCAATGGCACCATCAGGACGGGCTTC	1747
Qy	581	GlyGluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIle	600
Db	1748	GCGGAGAACCCGACACTTACGTGATGTGCCAGCTATCAAGAGCGCGAGGAGCATCATC	1807
Qy	601	AsnThrLeuLysGlyGlyThrAspGlyLysAsnThrGlyGluHisArgAsnLeu	618
Db	1808	AATACACTCAAGGTGGGACTGACGGGAAAAATACAGGCGGATCGCAACCTT	1861
RESULT 2			
AAZ46411			
ID AAZ46411 standard; DNA; 1902 BP.			
XX	AAZ46411;		
AC	AAZ46411;		
XX			
DT	07-MAR-2000 (first entry)		
XX			
DE	Pleurotus cornuopieae antitumour protein coding sequence.		
XX			
KW	Antitumour; cancer; tumour; treatment; expression; tumour suppressor;		
KW	p53; pBR; ss.		
XX			
OS	Pleurotus cornuopieae.		
XX			
PN	JP11315096-A.		
XX			
PD	16-NOV-1999.		
XX			
PF	07-AUG-1998; 98JP-0236349.		
XX			
PR	08-AUG-1997; 97JP-0215311.		
PR	02-MAR-1998; 98JP-0066176.		
XX			
PA	(NEWF-) NEW FOOD CREATION GIJUTSU KENKYU KUMIAI.		
XX			
DR	WPI: 2000-058170/05.		
DR	P-PSDB; AA52700.		
XX			
PT	An antitumour protein derived from Pleurotus cornuopieae and its gene -		
PT	useful for treatment of cancer including those caused by abnormal		
PT	expression of cancer inhibitory gene (e.g. p53 and pBR)		
XX			
PS	Claim 9; Page 15-16; 23pp; Japanese.		
XX			
CC	The invention relates to a novel antitumour protein extracted from		
CC	fruiting bodies of the fungus Pleurotus cornuopieae. The protein and		
CC	nucleotides encoding it are useful for the treatment of cancer, and		
CC	including those caused by abnormal expression of tumour suppressor		
CC	genes such as p53 and pBR. This sequence represents the coding		
CC	sequence of the gene encoding the antitumour protein.		
XX			
SQ	Sequence 1902 BP; 437 A; 556 C; 479 G; 430 T; 0 other;		
Alignment Scores:			
Pred. No.:	1.18e-98	Length:	1902
Score:	1232.00	Matches:	275

Percent Similarity:	57.05%	Conservative:	77
Best Local Similarity:	44.57%	Mismatches:	193
Query Match:	37.52%	Indels:	72
DB:	21	Gaps:	18
US-09-856-327-2 (1-618) x AAZ46411 (1-1902)			
Qy	18	IleAsnGlyGlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAsp	37
Db	130	ATTCGGGAAAACATTCCGTCGGGA-----GATACATCCAAGTTCGAC	171
Qy	38	valPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAla	57
Db	172	ETTGTCAATCGTCGGCTCTGGCCCAAGTTGGCTCTACTTATGCCGCTCTGCTGTCAGGCA	231
Qy	58	GlyLeuArgValValMetValGluThrIleGlyAlaAlaAspSerPheTyrAlaValAsnAla	77
Db	232	GTTTCAAAAGTAGGTATGTTGAATTGGAGAAATCGACTCT-----	273
Qy	78	GluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGluPhe	97
Db	274	-----GGCGCAAGCTT-----GGTTCCCATAGAATAATACGTTTTCAGTAC	315
Qy	98	GlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSerVal	117
Db	316	CAGAAGAACATTGCAAAATTCGTCCATGTCATTCAAGGACAACTCATGCTGTATCCCGTC	375
Qy	118	ProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTyrSerAlaProGly	137
Db	376	CCTGTCAACAAATATGTTGCGATACTCTCAGCCCGCTGAGTGGCAGGCC-----TCA	429
Qy	138	SerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAla	157
Db	430	ACACACTTGTCCCGCAAGCGCGCAACCCGGGACGAAGATCCCTTTACTTAATCTCTCAGGC	489
Qy	158	GluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArg	177
Db	490	CAAGCTGTACCCCGTGTGGGAGGAATGGCGACACTGGACTTGGCGAACACACTCGC	549
Qy	178	IleHisProProMetGluSerLeuProGlyIleGlyArgProLysLeu--SerAsnAsp	196
Db	550	TTCCACAAGAGT-----GAGCGGCCTTAAGCTTGTGAAGAGTATGAT	588
Qy	197	ProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGly	216
Db	589	GACTCCGGGACGCATCGATGGGACGCCCTCTATGACATCGCGAATGCTTCTGTCGAC	648
Qy	217	ThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGln	236
Db	649	ACCGGCATAATCAGTTTGATCAATCTATCCGGCACACCTTGTTACTCGAGAAGCTTCAG	708
Qy	237	AspAlaThrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeu	256
Db	709	GAGTCCCTTCAGCT---CAGCGCGGATTTGACGAGATCCCTCTTGGCGCCCAACGCACC	765
Qy	257	LysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPhe-----HisSer	274
Db	766	AAC-----CCCCCTTCGTCGATGGAGCTCTGCGCATACGGTCTTTGACCTTCAGAAC	819
Qy	275	IleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThr	294
Db	820	CGCCCAACCGGATGACGAGAAGGGCGTTTCAACCTCTTCCCGCGCATGTTGATGTGAA	879
Qy	295	ArgLeuAlaLeuThrGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeu	314
Db	880	CGGTTTACG---CGGGATTCCTCGACCGGAAGATAGAGGACATTGAGGTCCATGACCTG	936
Qy	315	LeuAlaThrArgAsnProSerSerGlnLeuAspSerTyr---IleMetAlaLysValTyr	333
Db	937	ATTAGCGGC-----GATCGCTACAAGGTCAAGGCTACAGCTGTTT	975
Qy	334	ValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGly	353

Db 976 ATCTTTTGTCTGCGCAGTCCATAATCCAGATTCTTGTAAACTCTGGATTC---GGG 1032
 QY 354 LeuGlnValThrProArgAsnAsp-----SerLeulleProAsnLeuGlyArg 369
 Db 1033 CGCATGGGTGAGCAGATTTCTCACTCCACCGCCGCTGCTGCGGTATCTCGGTAGC 1092
 QY 370 TyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAsp 389
 Db 1093 TACATCACTGAGCAGACACTCAGTCTGCGCAGACCGCTTTCAGCACCGAAGCTCGTCAAT 1152
 QY 390 SerValArgAspAspPro-----TyrGlyLeuPro----- 399
 Db 1153 CTGTGCAAGTCGACAGATGATTATTGTTGGCAGCCCGCAGACGACAGACTACAGTGTACT 1212
 QY 400 ----- 1212
 Db 1213 TTCACCCCGCAGACGTCGAGCAATAAGCACCAAAATGTTGGAACGAGAAGTCCAGAAG 1272
 QY 408 HisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGln 427
 Db 1273 CACATGATGACGACAGCAGGAGGATCCCTCCCGATCCCGTCGATGACCCCTGAGCCACAG 1332
 QY 428 ValThrThrProPheThrGluGlnHisProThrPheIleHisArgAspAlaPhe 447
 Db 1333 GTCACGACACTCTTCCAGGACACACATCCATGCGCACACTTCAGATTTCGTGATGCTTT 1392
 QY 448 SerTyrGlyAlaValGlyProGluValAspSerArgValIleValAspLeuArgTrpPhe 467
 Db 1393 AGCTACGGCGCTGTGGCGGAGATATCGACAGCGCTGTTGTTGACTGGCGCTCTTC 1452
 QY 468 GlyAlaThrAspProGluAlaAsnAsnLeuLeuValPheGlnAsnAspValGlnAspGly 487
 Db 1453 GGTGCGCACTGAACCTGTGGAGGAGACAAGCTGTGTTCTCAAAACAGACTGATGCA 1512
 QY 488 TyrSerMetProGlnProThrPheArgTyrArg---ProSerThrAlaSerAsnValArg 506
 Db 1513 TACAACTGCCACAGCCCACTTACAGCTTCCGTTCCCGGCGCCGCGCACAGCCAGGAG 1572
 QY 507 AlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGlyTyrLeuPro 526
 Db 1573 GCGGACCTCATGATGCTGACATGTGACGATGTCAACAAAGTCTGCTGCTCTGCTG 1632
 QY 527 ThrSerProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArg 546
 Db 1633 GGTTCATATCCGACGTTTATGCTCTGCGCTTGACTTCACTTGGTGGAAACCCACCGC 1692
 QY 547 IleGlyPheAspLysAlaThrThrValAla-----AspAsnAsnSerLeuValTrpAsp 564
 Db 1693 ATGGGCTTTGATGAGGAAGCTGATAAGCGTGTGTCGACACCAACTCAAAAGTCTTCGCG 1752
 QY 565 PheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnPro 584
 Db 1753 ATGGAGAACTGTCTCTGGCGGTTGCGGCAACATTTGGCAGCGGTATGCTGCTGCAACCG 1812
 QY 585 ThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsn 601
 Db 1813 ACGCTCACTGCGGTTCGCTTGCATCAGGAGCTGTAAAGTATATATATATATATATATAT 1863
 RESULT 3
 AAT34420
 ID AAT34420 standard; cdNA; 1869 bp.
 XX
 AC AAT34420;
 XX
 DT 27-NOV-1996 (first entry)
 XX
 DE Pyranose oxidase encoding sequence.
 XX
 KW Pyranose oxidase; glucose; oxidation; glucanase; assay; diabetes;
 KW marker; diagnosis; 1,5-anhydro-D-sorbitol; ss.
 XX
 OS Coriolus versicolor.
 XX

PN DEL9545780-AL.
 XX 13-JUN-1996.
 XX PF 07-DEC-1995; 95DE-1045780.
 XX 24-MAY-1995; 95JP-0124835.
 PR 07-DEC-1994; 94JP-0304086.
 XX (KIKK) KIKKOMAN CORP.
 XX Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K;
 PI Suzuki M;
 XX WPI; 1996-278990/29.
 DR P-PSDB; AAR99628.
 XX
 PT DNA encoding protein with pyranose oxidase activity at neutral pH -
 useful for the determination of glucose in body fluids or foods, or
 1,5-anhydro-D-sorbitol used as marker for diabetes diagnosis
 XX
 PS Claim 1; Page 10-13; 22pp; German.
 CC The present sequence encodes a protein isolated from Coriolus versicolor,
 which has the enzyme activity of pyranose oxidase (PO). The PO oxidises
 glucose to gluconone and has an optimum pH of 7-7.5. It has a mol. wt. of
 230000 (determined by gel filtration) and is stable at around 50deg.C.
 CC PO can be used for measurement of glucose in, e.g. foods or body fluids,
 or 1,5-anhydro-D-sorbitol which is an important marker used in the
 CC diagnosis of diabetes.
 XX
 SQ Sequence 1869 BP; 408 A; 601 C; 521 G; 339 T; 0 other;
 Alignment Scores:
 Pred. No.: 7 9e-98 Length: 1869
 Score: 1222.50 Matches: 272
 Percent Similarity: 56.84% Conservative: 73
 Best Local Similarity: 44.81% Mismatches: 177
 Query Match: 37.23% Indels: 85
 DB: 17 Gaps: 16
 US-09-856-327-2 (1-618) x AAT34420 (1-1869)
 QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuGlyValGlu 56
 Db 142 GACGTTGTATAGTAGCTCCGACCGATTGGATGACCTATGCGCGTGAGCTCGTAA 201
 QY 57 AlaGlyLeuArgValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
 Db 202 GCCGGTTACAGGTGCGCATGTTGCGACATCGGGGAAATGACTCTGGCTGAAGATC--- 258
 QY 77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGlu 96
 Db 259 -----GGTGCCCAAGAAAGAACCCGCTCGAA 285
 QY 97 PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnValSer 116
 Db 286 TACCAGAAGAACATTGACAAAGTTGTGAACGTCATTGAGGCCAATTGATGCTGTTCC 345
 QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaPro 136
 Db 346 GTTCCCGTCAATACCTCGTGTGATGACAGCGCTCAGCCCGACGCTTGTGCAAGCT--- 399
 QY 137 GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSer 156
 Db 400 TCATCGTCTTCCTCCGCAATGCTCGAACCAGACAGAGCCCGCTTCGTACCTCAGT 459
 QY 157 AlaGluAlaValThrArgGlyValGlyMetSerThrHisTrpThrCysSerThrPro 176
 Db 460 GGTACGGCGGTACACGCGTGTCTGCGGAGCATGTCCACGACTGCGACATGCGCACACCG 519
 QY 177 ArgIleHisProProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAsp 196
 |||

```
Db 520 CGTTTACC GCGGAGCAG-----CGCCCGTTGCTCGTGAAGAC 558
Qy 197 ProAlaGlu----AspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIle 215
Db 559 GACCAGGAGCTGACGAGCGGAGTGGGACCGGCTGTACACCAAGCGCGATCATCTTC 618
Qy 216 GlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeu 235
Db 619 AAGACCGCGGACGAGCGGAGTTCAGAGAGTCGATCCGCCACAACCTCGTCTCAACAAGTTC 678
Qy 236 GlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg 255
Db 679 GCGGAGGAAACAAGGT---CAGCGGACACTCCAGACAGATCCCCCTCGCGGCAACGCGT 735
Qy 256 LeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPhe-----His 273
Db 736 -----CGCAGCTCCGACCTTCGTCGAGTGGAGCTCGGGGAACACCGTGTTCGACCTCCAG 789
Qy 274 SerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCys 293
Db 790 AACAGCGCGAAGACGAGCGGACCGCAATGAGCGCTTCAACCTCTTCCCGCGGTTGCATGT 849
Qy 294 ThrArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsn 313
Db 850 GAGCGCGTC-----GTGCGCAAC 867
Qy 314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSertyr----- 327
Db 868 -----ACGTCAACTCCGAGATCCGAGAGTCTGCACATCCACGACCTCATC 912
Qy 328 -----IleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGly 341
Db 913 TCCGCGGACCGCTTCGAAATCAAGACGACGTGTTCGTCTTACAGCGGCGGCTCCAC 972
Qy 342 AsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeu---GlnValThrProArgAsn 360
Db 973 AACCGCGAGCTTCGTGAATCTCGGCTTTCGACAGCTGGCGCGCGGACCCCGCGAAC 1032
Qy 361 ---AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCys 379
Db 1033 CCGCGCGAGTTGCTCCGCTCCCTCGGAAGTATACATCCAGGACGAGTCTGCTCTTCGC 1092
Qy 380 GlnIleValLeuArgGlnGluPheValAspSerValArgAspAspProTyr-----Gly 397
Db 1093 CAGACCGTGATGAGCAGCGGAGCTCATCAGACGCTCAAGTCCGACATCATCAGGGGC 1152
Qy 398 LeuPro----- 399
Db 1153 AACCCCTGGCATCTGGGGTACAGCGTCAAGCTACAGCGCCCGCGGAGACCAACAAGCAC 1212
Qy 400 -----TrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
Db 1213 CCGGACTGGTGGAAACGAAGGTCAAGAACCATATGATGACAGCAGGAGGACCGCGTT 1272
Qy 418 ProIleProPheArgAspProGluProGlnValThrProPheThrGluGluHisPro 437
Db 1273 CCAATCCGTTCCGAGGACCGCGAGTCCAGCTCACCACCTTGTTCCAGCCATCCACCCG 1332
Qy 438 TrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAsp 457
Db 1333 TGGCACACTCAGATTACCGCGATGCGGTTCAGTTACGCGCGGTGTCAGCAACACCTGAC 1392
Qy 458 SerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAspLeu 477
Db 1393 TCAGCTCATCTGCTGACTGCGGTCTTTCGCGCGGAGGAGCCAAAGGAGGAGCAACAAG 1452
Qy 478 LeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrArgTyr 497
Db 1453 CTCTGGTTCGCGACAAAATTACGACACGCTTACCAATCGCGACGCGGCTTCGACTTC 1512
Qy 498 Arg---ProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGlu 516
Db 1513 CGTTCCCGGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1572
```

```
Qy 517 ValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProGlnPheMetAspProGly 536
Db 1573 ATGTCCGGGAGAAGATTGGTGGCTCTCTGCGCGCTCCCTCCCGCAATTTCATGAGCCCGGT 1632
Qy 537 LeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAla-----ThrThr 554
Db 1633 CTGTCTCTTCACTCCGTTGAGTACGACGCGCATGGGCTTCGACGAGCAGGAGCAAGTGC 1692
Qy 555 ValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGly 574
Db 1693 TCGGTCAACACGGGACTCCGCGTGTTCGCTTCAAGAACCCTGTCTCGTGGCTGCGGA 1752
Qy 575 ThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAlaIleLys 594
Db 1753 AACATCCACCGCGGTACGCGCGCAACCCGACGCTCACCGCAATGCTCGCTCGGATCAAG 1812
Qy 595 SerAlaArgSerIleIleAsn 601
Db 1813 AGTTGCGAGTACATCAAGAAC 1833
RESULT 4
AAC87519
ID AAC87519 standard; DNA; 1869 BP.
AC AAC87519;
XX
AC AAC87519;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trametes hirsuta pyranose oxidase cDNA, SEQ ID NO:1 (version 2).
XX
KW Pyranose oxidase; expression construct; recombinant production;
KW monosaccharide oxidation; 2-keto derivative;
KW hydrogen peroxide production; ss.
XX
OS Trametes hirsuta.
XX
PN US6146865-A.
XX
PD 14-NOV-2000.
XX
PF 05-MAY-1999; 99US-0305381.
XX
PR 08-JUN-1998; 98DK-0000774.
XX
PR 10-JUN-1998; 98US-0088724.
XX
PA (NOVO ) NOVO NORDISK AS.
XX
PI Schneider P, Christensen S, Lassen SF;
XX
WP1; 2001-049055/06.
XX
P-PSDB; AAB48832.
XX
PT Novel nucleic acid molecule encoding polypeptide having pyranose
PT oxidase activity used to design oligonucleotide probes to identify and
PT clone DNA encoding the polypeptide from different genera or species -
XX
XX
PS Claim 2; Column 25-28; 20pp; English.
XX
XX
CC The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which
CC encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also
CC relates to expression constructs, expression vectors and recombinant
CC cells comprising pyranose oxidase nucleic acid sequences, and the
CC recombinant production of Trametes hirsuta pyranose oxidase. Pyranose
CC oxidase catalyses the oxidation of several monosaccharides in the
CC pyranose form at position C2 to produce 2-keto derivatives with the
CC release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta
CC pyranose oxidase may be used to produce the enzyme and to design
CC oligonucleotide probes to identify and clone genomic pyranose oxidase
CC cDNA or genomic DNA from different genera or species of microorganisms
CC (fungi or bacteria). The present sequence represents a cDNA encoding
CC pyranose oxidase from the fungus Trametes hirsuta.
XX
Note: Both AAC87518 and AAC87519 are Trametes hirsuta pyranose oxidase
```


QY 426 ProGlnValThrProPheThrGluGluHisProThrHisThrGlnIleHisArgAsp 445
 Db 1329 CCGCAGGTCAACCGCTGTTTCAGGCAACGACGCCACCCAGATTCACCCGAC 1388
 QY 446 AlaPheSerTyrglyAlaValGlyProGluValAspSerArgValIleValAspLeuArg 465
 Db 1389 GCCTTCAGTACGGCCGCTGCAGCAGCATCGACTCGGGCTCATCTGCTCGACTGGCGG 1448
 QY 466 TrpPheGlyAlaThrAspProGluAlaAsnLeuValPheGlnAsnAspValGln 485
 Db 1449 TTCTTCGGACGACCCAGCCAGGAGGAGAGCAAGCTATGTTCTCGCAGACATCACG 1508
 QY 486 AspGlyTyrrSerMetProGlnProThrPheArgTyrrArgProSerThrAlaSerAsnVal 505
 Db 1509 GACGCTACAACTTCGCGCAGCCGACGTTTCGACTTCGCGC-----TTCGCCGGGGCCGC 1562
 QY 506 ArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrrLeu 525
 Db 1563 GAAGCGGAGCATATGATGACCATGTCGTCATGTCGCGGAGATCGGTGATTCCTG 1622
 QY 526 ProThrSerProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThr 545
 Db 1623 CCTGGTCTCAACCCAGTTCATGGAGCCGCTGTTGCTCTGCACCTTGTGGGACGCAC 1682
 QY 546 ArgIleGlyPheAspLysAlaThr-----ThrValAlaAspAsnAsnSerLeuValTrp 563
 Db 1683 CGCATGGGCTTCGACGAGAGCGGACAAAGTGTGCTGTCGACACCGACTCAACGCGTCTTC 1742
 QY 564 AspPheAlaAsnLeuTyrrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsn 583
 Db 1743 GCGTTCAGAACTCTTCCTCGCGGCTCGCGGAACATCCCAACCGGTACCGCGCGGAAC 1802
 QY 584 ProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsnThrLeu 603
 Db 1803 CCGACGCTCACCGCAATGCTGCTTCGATCAAGAGCTGCGAGTACATC-----AAG 1853
 QY 604 LysGlyGlyThrAspGlyLysAsnThrGlyGluHisArgAsn 617
 Db 1854 AAGAACTTCGACCGCGAGCCGCAACCCGCGTGAAGCACCACAAC 1895
 RESULT 6
 AAV83626
 ID AAV83626 standard; cDNA to mRNA; 1701 BP.
 XX
 AC AAV83626;
 XX
 XX
 DT 26-FEB-1999 (first entry)
 XX
 DE Nucleic acid encoding an antitumour protein.
 XX
 KW Antitumour protein; Tricholoma matsutake; ss.
 XX
 XX Tricholoma matsutake.
 OS
 XX
 XX
 FT Key Location/Qualifiers
 CDS 1..1701
 FT /*tag= a
 FT /product= antitumour_protein
 XX
 PN JP10313876-A.
 XX
 PD 02-DEC-1998.
 XX
 PF 13-FEB-1998; 98JP-0031452.
 XX
 PR 13-FEB-1997; 97JP-0029275.
 XX
 PR (MOMO-) MOMOYA KK.
 PA (NORQ) NORINSUISANSO SHOKUJIN SOGO.
 XX
 WIPI; 1999-074153/07.
 DR P-PSDB; AAW87531.
 XX

PT An anti-tumour protein - prepared by culture of host cell
 PT transformed by vector containing base coding sequence
 XX
 PS Cladm 3; Page 8-10; 15pp; Japanese.
 XX
 CC The present sequence encodes an antitumour protein, and is isolated
 CC from Tricholoma matsutake.
 XX
 SQ Sequence 1701 BP; 411 A; 463 C; 432 G; 395 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,68e-86 Length: 1701
 Score: 1091.00 Matches: 248
 Percent Similarity: 57.04% Conservative: 84
 Best Local Similarity: 42.61% Mismatches: 184
 Query Match: 33.22% Indels: 66
 DB: 20 Gaps: 15

US-09-856-327-2 (1-618) x AAV83626 (1-1701)

QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrrAlaLysLeuCysValGlu 56
 Db 97 GATGTTTTCATTCCTGGCAGTGGTCCCATTAAGTCTACTTACGCCGCCCATCATGTGAC 156
 QY 57 -----AlaGlyLeuArgValMetValGluIleGlyAlaAlaAspSerPheTyrrAla 74
 Db 157 AATACCTCAACTACAAAGGTTTACATGCGCGCAATAGGTTCTCAAGATAAC----- 207
 QY 75 ValAsnAlaGluGluGlyThrAlaValProTyrrValProGlyTyrrHisLysLysAsnGlu 94
 Db 208 -----CCTGTCATC-----GGGCGCCCATCACAGAACTCC 237
 QY 95 IleGluPheGlnLysAspIleAspPheValAsnValIleLysGlyAlaLeuGlnGln 114
 Db 238 ATAAAGTTTCAAGAGACATTCACAAGTTTGAATATCATCAACGTCCTCCACGCG 297
 QY 115 ValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAla 134
 Db 298 ATTTCGATTTCCGCATCGGACACCTACACAGCCACTCTCGCTGTAGACGCTGG---GCG 354
 QY 135 ProProGlySerSerAla-----IleSerAsnGlyLysAsnProHisGlnArg 150
 Db 355 CCGCCCATCGATCTCGCCGAAGCCAGCTCGGTATATGGACACATCCGAATCAGGAG 414
 QY 151 GluPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyMetSerThrHis 170
 Db 415 GCGGCGCTGAACCTCCCGGTAGCGCTGTCACTAGGACAGTCGGGGGATGGCACCAC 474
 QY 171 TrpThrCysSerThrProArgIleHisProMetGluSerLeuProGlyIleGlyArg 190
 Db 475 TGGACTTGGCGGTGCTCTACTCCACATGAC----- 504
 QY 191 ProLysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyrrSerGlu 210
 Db 505 GAAGAGAGGTTCAACAACCCAGTT---GACAGAGGAGTTCGACACATGCTCGAAGCT 561
 QY 211 AlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeu 230
 Db 562 GCTAAACATTCCTCAACGTTCCACGCGACAGTAGGACGATCTATCTCGTCAGATAGTT 621
 QY 231 ValLeuArgSerLeuGlnAspAlaTyrrLysAspArgGlnArgIlePheArgProLeuPro 250
 Db 622 GTCAAAGAGACTCTTCAGCAGACCCCTT---GATGCGTCGCGGGGTGTGACCACTCTCCCG 678
 QY 251 LeuAlaCysHisArgLeuLysAsnAlaProGluTyrrValGluTrpHisSerAlaGluAsn 270
 Db 679 CTGGGGGTGGAGCGCGGTACGACCAATCTATTATGTCACCTGGACCGGTCGCGCAT--- 735
 QY 271 LeuPheHisSerIleTyrrAsnAspLysGlnLysLysLeuPheThrLeuThrAsn 290
 Db 736 -----ACCGTCTTGGTGTATGCGGAAGAGTCCCGGATTCGCTTGTGTACAGAG 786
 QY 291 HisArgCysThrArgLeuAlaLeuThrGlyGlyTyrrGlyLysLysIleGlyAlaGlu 310


```
Db 787 ACAGAGTGCAGGAGCTTATTGTCAGTGAACCAATCGACGAGGTGTGTCGCGTTG 846
Qy ValArgAsnLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAla 330
Db 847 CTACGTAACCTG-----AATCAAGCAAC-----GATGAACCTGTGCTGGCC 888
Qy LysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGly 350
Db 889 AAGATTTCGTCATAGTGTGGAGCAGCTGCACACCGCAAACTTTGTGGAACAGC--- 945
Qy 351 PheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyr 370
Db 946 -----AACATCCGCCCATAT-----GCGCTTGGTGGCTGAC 975
Qy 371 IleThrGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSer 390
Db 976 CTCAGCAGNACAGTCCATGACCTTTTGTGCAGATCTTCTCAAGAGGGCATAGTCATGCC 1035
Qy 391 ValArgAspAspProTyrGlyLeuProTyrTrpLysGluAlaValAlaGlnHisIleAla 410
Db 1036 ATCGCTACTGACCTCGC-----TTCGCTGCGAAGGTTGAGGCGCACAGAAG 1083
Qy 411 LysAsnProThrAspAlaLeuProIleProPheArgaspProGluProGlnValThrThr 430
Db 1084 AAGCACCCTGATGAGCTGCTGCCATTCATTCACGAGCCTGAACCTCAAGTATGATGATT 1143
Qy 431 ProPheThrGluGluHisProTyrHisThrGlnIleHisArgAspAlaPheSerTyrGly 450
Db 1144 CCGTACAGCTCGGACTCCCTTGGCATGTTCCAGTGCACTCCGATGCTCTCATATGGT 1203
Qy 451 AlaValGlyProGluValAspSerArgValIleValAspLeuArgTyrPheGlyAlaThr 470
Db 1204 GATGTTGGACCAAGCGGACCGCGCTGTTGTCGTCGATCTGAGGTTTTTCGGCAAAATCA 1263
Qy 471 AspProGluAlaAsnAsnLeuLeuValPhe-----Gln 481
Db 1264 GATATTGTCGAAGAAATCGAGTCACTTTCGCTCGCAACCCCTAAGCTACGCGAGTGGGAA 1323
Qy 482 AsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThr 501
Db 1324 CGCGGTGTACAGACACTATGGAATGCCAGCGGACATTCATGTCGAACGGACCAAC 1383
Qy 502 AlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeu 521
Db 1384 GCCGATGGAGACCGTGACCGAGGATGATGAATGATATGACCAACGTCGCGCAACATGCTG 1443
Qy 522 GlyGlyTyrLeuProThrSerProGlnPheMetAspProGlyLeuAlaLeuHisLeu 541
Db 1444 GGTGGGTACCTTCTGCTGCTTACCTCAATTTATGGCACCCTGCTGCTGCTACTGCACATC 1503
Qy 542 AlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeu 561
Db 1504 ACGGGAACACTACTCGATCGGACAGATGATCAAACTCTGTGTCATGCCACATCAAG 1563
Qy 562 ValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGly 581
Db 1564 GTTCATAACTTCAACAATCTGTGGTCGGCGGAATGGTGCTCATTCAGATGCGACTGCC 1623
Qy 582 GluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsn 601
Db 1624 TGCACCCGACTCTGACGAGCGTCGCTATCGCTCAAGGGTGTGAGGCTGTATGATCAAT 1683
Qy 602 ThrLeu 603
Db 1684 TACCTT 1689
```

RESULT 7

AAA71487

ID AAA71487 standard; cDNA; 1946 BP.

XX

AC AAA71487;

XX

```
DT 11-DEC-2000 (first entry)
XX
DE T. matsutake pyranose oxidase cDNA.
XX
KW Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
KW diabetes mellitus; ss.
XX
OS Tricholoma matsutake.
XX
FH Key Location/Qualifiers
FT CDS 101..1795
FT /*tag= a
FT /product= "pyranose oxidase"
XX
PN JP2000175698-A.
XX
PD 27-JUN-2000.
XX
PF 16-DEC-1998; 98JP-0357423.
XX
PR 16-DEC-1998; 98JP-0357423.
XX
PA (NIBS ) JAPAN TOBACCO INC.
XX
DR WPI; 2000-478485/42.
DR P-PSDB; AAB10457.
XX
XX A reagent containing pyranose oxidase for the determination of pyranose
PS Disclosure; Page 14-17; 23pp; Japanese.
XX
CC This invention describes a novel reagent containing pyranose oxidase for
CC the determination of pyranose which can be prepared from a fraction
CC precipitated from an aqueous extract of matsutake mushroom by ammonium
CC sulfate precipitation and has an antibacterial activity against at least
CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
CC 210 kD by gel filtration and shows the presence of components of ca.
CC 50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
CC activity by being heated at 60 degrees C for 10 minutes in a neutral
CC aqueous solution and in which the above antibacterial activity is
CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
CC aqueous solution. The invention also describes a method for the
CC determination of pyranose in a sample in which the above reagent for the
CC determination of pyranose is reacted with pyranose in the sample and the
CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
CC method for diagnosing a disease accompanied by abnormality in sugar
CC metabolism in which the above reagent for the determination of pyranose
CC is reacted with an enzyme to develop a color, and a kit for pyranose
CC analysis or the diagnosis of a disease accompanied by abnormality in
CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
CC for diabetes mellitus. This sequence encodes the Tricholoma matsutake
CC pyranose oxidase protein which is described in the method of the
CC invention.
SQ Sequence 1946 BP; 484 A; 516 C; 475 G; 471 T; 0 other;
```

Alignment Scores:

Pred. No.:	4,48e-82	Length:	1946
Score:	104.00	Matches:	241
Percent Similarity:	55.91%	Conservative:	90
Best Local Similarity:	40.71%	Mismatches:	187
Query Match:	31.79%	Indels:	74
DB:	21	Gaps:	17

US-09-856-327-2 (1-618) x AAA71487 (1-1946)

Qy 27 IleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyProfile 46

Db 185 GTACATTACACT-----GATGTTTCATTCGTCGACGATGGTCCCAT 226

Qy 47 GlyAlaThrTyrAlaLysLeuCysValGlu-----AlaGlyLeuArgValMetVal 64

```
Db 227 GCGTGTACTACGCGCGGCACATCATTTGACAATACTCAACTACAAAGGCTCTACATGGCC 286
QY 65 GlnIleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGluGlyThrAlaValPro 84
Db 287 GAATAGGTTCTCAGATAAC-----CCT 310
QY 85 TyrValProGlyTyrHisLysLysAsnGluIleGluPheGlnLysAspIleAspArgPhe 104
Db 311 GTCATC---GGAGCCCATCACAGGAACCTCCATAAAGTTTCAGAAAGACACTCACAAGTTT 367
QY 105 ValAsnValIleGlyAlaLeuGlnGlnValSerValProValArgAsnGlnAsnVal 124
Db 368 GTGAATATCATCAACGGTGCCTCCAGGCCATTTCGATTTCGCCATCGGACACCTACCAG 427
QY 125 ProThrLeuAspProGlyValArgPheSerAlaProGlySerSerAla----- 140
Db 428 CCCACTCTCGCTGTAGCAGCGTGG---GCGCCGCCCATCATGATCTCGCAAGGCCAGCTC 484
QY 141 IleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
Db 485 GTGATTATGGACACAATCCGAATCAGGAGGCCGCGCTGAACCTTCCCGGTAGCGCTGTC 544
QY 161 ThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHisPro 180
Db 545 ACGAGGACAGTCGGGGGAATGGCGACCCACTGGACTTGGCGTGTCTACTCCACATGAC 604
QY 181 ProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAsp 200
Db 605 -----GAAGAGAGGGTCAACAACCCAGTT---GAC 631
QY 201 AspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLys 220
Db 632 AAGCAGGAGTTCGACGCACTGCTCGAACGTGTAAACATTCCTCAACGTTTCACAGCGAC 691
QY 221 GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys 240
Db 692 CAGTATGACGATTCTATCCGTGAGTAGTTGTCAAGAGAGACCCCTTCAGCAGACCCCTT--- 748
QY 241 AspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaPro 260
Db 749 GATGCGTCGCGGGTGTACCACTCTCCGCTGGGGTGGAGCGCGCGCGCAACATCTCT 808
QY 261 GluTyrValGluTrpHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLys 280
Db 809 ATTTATGTCACCTGGACCGGTCCGAT-----ACCGTCTTGGTGTGATGTGCGG 856
QY 281 GlnLysLysLeuPheThrLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly 300
Db 857 AAGATCCCGGATTCGTTTGGTTACAGAGAGAGAGTGCAGCAAGTTTATTGTCAGTGAA 916
QY 301 GlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 320
Db 917 ACCAATCCGACGCGGTTGTCGCGTGTGTACGTAACCTTG-----AATACA 964
QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
Db 965 AGCAAC-----GATGAATCTGCTGCGCCAGAGCTTCGTCATAGTTGTGGACGAGTC 1018
QY 341 GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn 360
Db 1019 TGCACACCGCAATCTGTGAAACAGC-----AACATCCGCCCCACAT--- 1060
QY 361 AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln 380
Db 1061 -----GCGCTGGTGGCTACCTACCGCAACAGTCCATGACTCTTTTGTGTGAG 1105
QY 381 IleValLeuArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProTrp 400
Db 1106 ATTGTTCTCAAGAGGAGCATAGTCGATTCCATCGCTACTGACCCCTCGC----- 1153
QY 401 TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro 420
Db 420 -----GAC 420
```

```
1154 TTCGTGCGAAGGTTGAGGCGCACAGAAGAACGCCGATGACGTGTGCGCATTCOA 1213
QY 421 PheArgAspProGluProGlnValThrThrProPheThrGluGluHisProTrpHisThr 440
Db 1214 TTCACGAGCCTGAACCTCAAGTGATTCGTACACAGTCGCGATTCCTTGGCATGTT 1273
QY 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgVal 460
Db 1274 CAGGTCCATCGC-----TATGCATTGGTGATGTGGACCAAGCCGACCGCGTGT 1327
QY 461 IleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuLeuValPhe 480
Db 1328 GTCGTGATCTGAGGTTTTTCGCAATCAGATATTGTGAAAGAAATCGAGTGACTTTC 1387
QY 481 -----GlnAsnAspValGlnAspGlyTyrSerMetPro 491
Db 1388 GGTCCGAACCCCTAAGCTACGCGACTGGGAAGCGGGTGTACAGACACTTATGGAATGCCA 1447
QY 492 GlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMet 511
Db 1448 CAGCCGACATTCCATGTCAAGCGGACCAACGCCGATGGAGACCGTGACCAAGGATGATG 1507
QY 512 AlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProGln 531
Db 1508 AATGATATGACCAACGTCGCAACACTACTGGCGGGTACTTCTGGCTCCTACCCCTCAA 1567
QY 532 PheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrArgIleGlyPheAspLys 551
Db 1568 TTTATGGCACCTGGTCTCGCACAGCACATCAGCGGAACACTCTCGGATCGGACAGATGAT 1627
QY 552 AlaThrThrValAlaAlaAspAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla 571
Db 1628 CAAACTTCTGTGTGATCCGACATCAAAGGTTTCAACTTCGACAATCTGTGGTCTGCGC 1687
QY 572 GlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
Db 1688 GGAATGGGTGATTCCAGATCGCACTGCCTCAACCCGACTCTGACGAGCGTGCATAT 1747
QY 592 AlaIleLysSerAlaArgSerIleIleAsnThrLeu 603
Db 1748 GCGCTTAAGGTGCTGAGGCTGTAGCTGTTACCTT 1783

RESULT 8
AAA07403
ID AAA07403 standard; DNA; 1946 BP.
XX
AC AAA07403;
XX
DT 07-JUL-2000 (first entry)
XX
DE Trichoderma derived antifungal protein coding sequence.
XX
KW Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
KW growth inhibitor; plant pathogenic fungi; antibacterial agent;
KW N-terminal fragment; ss.
XX
Trichoderma matsutake.
XX
WO200014242-A1.
XX
PD 16-MAR-2000.
XX
PF 19-AUG-1999; 99WO-JP04441.
XX
PR 08-SEP-1998; 98JP-0270606.
XX
PA (NTSB ) JAPAN TOBACCO INC.
PA (NORQ ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
PI Takakura Y, Kuwata S, Ohta S;
XX
DR WPI; 2000-256990/22.
DR P-PSDB; AAY81952.
```

XX Mushroom-derived antibacterial protein against plant pathogenic fungi
 PT of rice, with activity and thermal stability, obtainable cheaply on
 TT large scale, useful in agriculture
 XX

PS Claim 13; Page 41-45; 52pp; Japanese.

XX This sequence encodes the Trichoderma matsutake antibacterial protein of
 CC the invention. The protein has activity against at least Pyricularia
 CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an
 CC aqueous extract of a mushroom precipitated by the ammonium sulphate
 CC precipitation method. The protein has a molecular weight of about 210 kD
 CC as determined by the gel filtration method, includes components of about
 CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous
 CC neutral solution at 60 degrees C for 10 minutes but with loss of
 CC antibacterial activity after heating in the solution at 80 degrees C for
 CC 10 minutes. The protein is used for inhibiting the growth of plant
 CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
 CC It is useful in treating rice plants, and is applicable in agriculture as
 CC an antibacterial agent. The protein has activity at relatively low
 CC concentrations, and can be produced at low cost on large scale.
 XX

SQ Sequence 1946 BP; 484 A; 516 C; 475 G; 471 T; 0 other;

Alignment Scores:

Pred. NO.: 4.48e-82 Length: 1946
 Score: 1044.00 Matches: 241
 Percent Similarity: 55.91% Conservative: 90
 Best Local Similarity: 40.71% Mismatches: 187
 Query Match: 31.79% Indels: 74
 DB: 21 Gaps: 17

US-09-856-327-2 (1-618) x AAA07403 (1-1946)

Qy 27 IleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyProIle 46
 Db 185 GTACATTACAT-----GATGTTTTTCATTTGTCGAGTGGTCCCAT 226
 Qy 47 GlyAlaThrTyrAlaLysLeuCysValGlu-----AlaGlyLeuArgValValMetVal 64
 Db 227 GCCTGTACTACGCCGCCCATCATTTGACAAATACCTCACTACAAGGTCATACATGCC 286
 Qy 65 GlutIleGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGlyThrAlaValPro 84
 Db 287 GAAATAGGTTCTCAAGATAAC-----CCT 310
 Qy 85 TyrValProGlyTyrHisLysLysAsncluleGluPheGlnLysAspIleAspArgPhe 104
 Db 311 GTCATC---GGAGCCCATACAGGACTCCATAAGTTTCAGAAAGACACTGCAAGTTT 367
 Qy 105 ValAsnValIleLysGlyAlaLeuGlnGlnValSerValProValArgAsnGlnAsnVal 124
 Db 368 GTGAATATCATCAACGGTCCCTCCAGCCCATTTTCGATTCGCCATCGGACACCTACCAG 427
 Qy 125 ProThrLeuAspProGlyAlaTrpSerAlaProGlySerSerAla----- 140
 Db 428 CCCACTCTCGGTAGCAGCGTGG---GCGCCGCCCATCGATCTCTCGCAGAGGCCAGCTC 484
 Qy 141 IleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
 Db 485 GTGATTATGGGACAAATCCGAATCAGGAGCCGCCCTGGAACCTTCCCGGTAGCGCTGTC 544
 Qy 161 ThrArgGlyValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIleHisPro 180
 Db 545 ACGAGGACAGTCGGGGGAGATGGCCACCCACTGGAATTCGGCGTCTCTACTCCACATGAC 604
 Qy 181 ProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAspProAlaGluAsp 200
 Db 605 -----GAAGAGAGGGTCAACACCCAGTT---GAC 631
 Qy 201 AspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLys 220
 Db 632 AAGCAGGAGTTTCGACGCACTGCTCGAAAGCTGCTAAACATTTGCTCAACAGTTTCACAGGAC 691

Qy 221 GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys 240
 Db 692 CAGTATGACGATTCTATCCGTCAGATAGTTGTCRAAGAGACCTTTCACGACAGCCCTT--- 748
 Qy 241 AspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaPro 260
 Db 749 GATCGTCGCGGGTGTGACCACTCTCCCGTGGGGTGGAGCGCCGACGACAAATCCT 808
 Qy 261 GluTyrValGluTyrHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLys 280
 Db 809 ATTTATGTCACCTGACCGGTGCCGAT-----ACCGTCCTTGGTGTGATGTGCGG 856
 Qy 281 GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuLeuGly 300
 Db 857 AAGAGTCCCGCATTCGTTTGTGTTACAGAGAGAGAGTACGAAGTTTATTTGTCAGTGAA 916
 Qy 301 GlyTyrGluLysLysIleGlyAlaGluValArgAsnLeuLeuAlaThrArgAsnPro 320
 Db 917 ACCAATCCGACGCGAGGTGTTGCTCGGTGCTACGTAACTTG-----AATACA 964
 Qy 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
 Db 965 AGCAAC-----GATGAACCTGTGTCGCCCCGAGTTTCGTGTCATAGCTTGTGGACGATC 1018
 Qy 341 GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn 360
 Db 1019 TGCACACGCGCAATCCTGTGGAACAGC-----AACATCCGCCCATCAT--- 1060
 Qy 361 AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln 380
 Db 1061 -----GCGCTTGGTGGCTACCTCAGCGAACATGCTCATGACTTCTTTGTGTCAG 1105
 Qy 381 IleValLeuArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProIrp 400
 Db 1106 ATTGTTCTCAAGAGAGCATAGTCGATTCCATCGCTACTGACCCCTCGC----- 1153
 Qy 401 TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro 420
 Db 1154 TTCGCTCGGAAGTTGAGCGCCACAAGAAGAACCCCGATGAGTGTGCTCCGATTCCA 1213
 Qy 421 PheArgAspProGluProGlnValThrThrProPheThrGluGluHisProThrIleHisThr 440
 Db 1214 TTCACGACGCTGAACCTCAAGTATGATGATTCGTCACACGTGGACTTCCCTTGGCATGTT 1273
 Qy 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAspSerArgVal 460
 Db 1274 CAGGTCCATCCG-----TATGCATTTGGTGTGATGTGGACCCCAAGCCGCGCGTGT 1327
 Qy 461 IleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeuValPhe 480
 Db 1328 GTCGTCGATCTGAGGTTTTTCGCAAAATCAGATATTTGTCGAAGAAAATCGAGTGACTTTC 1387
 Qy 481 -----GlnAsnAspValGlnAspGlyTyrSerMetPro 491
 Db 1388 GGTCGCAACCCCTAAGACTACGCGACTGGGAAGCGGTGTACAGACACTTATGGAATGCCA 1447
 Qy 492 GlnProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMet 511
 Db 1448 CAGCGCATATCCATGTCAGCGGACCAACCCGATGGAGACCGGTGACGAGGATGATG 1507
 Qy 512 AlaAspMetCysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProProGln 531
 Db 1508 AATGATATGACCAACGTCGCGCAACATACTGGGCGGTACTCTCTGGCTCCTACCTCAA 1567
 Qy 532 PheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrArgIleGlyPheAspLys 551
 Db 1568 TTTATGGCACCTGGTCTCGCACAGCATACCGGAACACTACTCGGATCGGACAGATGAT 1627
 Qy 552 AlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAla 571
 Db 1628 CAAACTCTGTGTGATCCGACATCAAAAGGTTTCATAACTTCGCAAAATCTGTGGTCCGCG 1687

```
QY 572 GlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
Db 1688 GGGATGGGTGCTCCAGATCGGACTCGCTGCAACCGAGCTCGTACGAGCGTCGCGTAT 1747

QY 592 AlalIleLysSerAlaArgSerIleIleAsnThrLeu 603
Db 1748 GCGCTTAAGGGTGTGAGGCTGTAGTACGTACCTT 1783

RESULT 9
ID ABK52367 standard; DNA; 2467 BP.
XX ABK52367;
XX
XX 27-AUG-2002 (first entry)
XX
XX DNA encoding glucose dehydrogenase associated proteins.
XX
XX Glucose dehydrogenase; electrode; glucose sensor; glucose-assay;
KW substrate specificity; heat stability; gene; ds.
XX
XX Burkholderia cepacia.
XX
XX Key Location/Qualifiers
FH CDS 258..764
FT /*tag= a
FT /product= "Glucose dehydrogenase associated protein #1"
CDS 764..2383
FT /*tag= b
FT /product= "Glucose dehydrogenase associated protein #2"
CDS 2386..2466
FT /*tag= c
FT /product= "Glucose dehydrogenase associated protein #3"
XX
XX WO200236779-A1.
XX
XX 10-MAY-2002.
XX
XX 31-OCT-2001; 2001WO-JP09556.
XX
XX 31-OCT-2000; 2000JP-0332085.
XX
XX 24-NOV-2000; 2000JP-0357102.
XX
XX 12-SEP-2001; 2001JP-0276832.
XX
XX (SODE/) SODE K.
XX
XX Sode K;
XX
XX WPI; 2002-463413/49.
XX
XX P-PSDB; AAU97825, AAU97826, AAU97827.
XX
XX production of Burkholderia glucose dehydrogenase for use in glucose
XX sensor electrodes and glucose-assay kits in medicine, science and
XX industry -
XX
XX Claim 17; Page 51-55; 61pp; Japanese.
XX
XX The invention describes a method of producing a glucose dehydrogenase
XX comprising culturing a Burkholderia microorganism and collecting the
XX product from the medium and/or the microbial cells. Glucose dehydrogenase
XX is useful in electrodes of glucose sensors and glucose-assay kits for
XX medicine, science and industry. Glucose dehydrogenase is economically
XX produced with high substrate specificity and improved heat stability to
XX provide long-term accuracy. This sequence encodes 3 glucose dehydrogenase
XX associated protein described in the invention.
XX
XX Sequence 2467 BP; 497 A; 792 C; 769 G; 409 T; 0 other;

Alignment Scores:
Pred. No.: 2,61e-08 Length: 2467
Score: 208.50 Matches: 131
Percent Similarity: 34.05% Conservative: 91
Best Local Similarity: 20.09% Mismatches: 221
```

```
Query Match: 6.35% Indels: 209
DB: 24 Gaps: 29
US-09-856-327-2 (1-618) x ABK52367 (1-2467)
QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
Db 791 GACGTGCTGCTGGTGGATCGGTCGCGGCGGATCGTGCAGTCAGTCGCGATG 850
QY 57 AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
Db 851 GCGGGCAAGCGGTGATCTCTCGAAGCGGCG----- 883
QY 77 AlaGluGluGlyThrAlaValProGlyTyrHisLysLysAsnGluIleGlu 96
Db 884 -----CCGCGCATGCCGCGCTGG-----GAAATC--- 907
QY 97 PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnValSer 116
Db 908 -----GTGAGCGCTTCGCAAT----- 925
QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSer----- 133
Db 926 CAGCCCGACAAAGATGGACTTCATGGCGCGCTACCGTCGAGCCCTGGGCGCGCATCC 985
QY 134 -----AlaProProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGlu 151
Db 986 GAGTACGCGCGCGCAGCACTACCTGATCTCGAAGGCGGAGCACAAGTTCACATCGCAG 1045
QY 152 PheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrp 171
Db 1046 TAC-----ATCGCGCGGTGGCGGCGCAGCAGCTGGCAGCTGG 1081
QY 172 ThrCysSerThrProArgIleHisPro-----PrometGluSerLeuProGlyIle 188
Db 1082 GCCGCGTCGGCGTGGCGCTTCCGAAACGACTTCAAGATGAAGAGCGGTACGCGCGTC 1141
QY 189 GlyArgProLysLeuSerAsnAspProAlaGlu---AspAspLysGluTrpAsnGluLeu 207
Db 1142 GCGCGC-----GACTGGCGGATCCAGTACGACGATCTCGAG-----CCGTAC 1183
QY 208 TyrSerGluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArg 227
Db 1184 TATCAGCGCGGAGGAGAGCTCGCGTGTGGCGCGCGCGCGCGAGAA----- 1234
QY 228 HisThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePhe--- 246
Db 1235 -----GATCTGTACTCGCGCGCAAGCAGCGCGGTATCCG 1267
QY 247 ---ArgProLeuProLeuAlaCysHisArg-----LeuLysAsnAlaProGluTyrVal 263
Db 1268 ATCGCGCGCGTCCGCTTCTGTTCAACGAGCAGACCATCAAGACGCG----- 1315
QY 264 GluTrpHisSerAlaGluAsnLeuPheHisSerIle-----TyrAsnAsp 278
Db 1316 ---CTGAACAACACTACGATCCGAAAGTTCCATGTCGTGACCGCGCGCGCGCAACAGC 1372
QY 279 AspLysGlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeu 298
Db 1373 CGCGCGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1432
QY 299 ThrGlyGlyTyr-----GluLysLysIleGlyAlaAla 309
Db 1433 ATCGCGCGGATGTACAACCGCATCGTCGACGTCGAGAGCGCGCAACGCGCGCGCGCG 1492
QY 310 GluValArgAsnLeuLeuAlaThrArg-----AsnProSerSerGlnLeu----- 324
Db 1493 CTGATCGAAGACCGCGGTCTGTACAGCTCGAGACGCGCGCGCGCGCGCGCGCGCG 1552
QY 325 -----AspSerTyrIleMetAlaLysValTyrValLeu 335
Db 1553 GCGCTCTACAAGGACAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1612
```


QY 81 ThrAlaValProTyrValProGlyTyrHisLysLysAsnGluLeuGluPheGlnLysAsp 100
 Db 1065 -----CGTATGATCGCGCGAGGCC 1085
 QY 101 IleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSerValProValArg 120
 Db 1086 GTGGAAACTGGCGCAAT-----ATGCGGTTCGCA 1115
 QY 121 AsnGln-----AsnValProThrLeuAspProGlyAlaTrpSerAlaPro----- 135
 Db 1116 AACCGCGCAGGCTCCGACTTTCAGGATATATCTCAGTCAAAATTTGCTCCGCGCAGC 1175
 QY 136 -----ProGlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgLuphe 152
 Db 1176 CTCATATTTCCCGCTAACTATGTC-----AACGTGACGGGTCCCAATGCTGACAGTTT 1232
 QY 153 GluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHisTrpThr 172
 Db 1233 CAA-----CAAGCTACCTGCGCAGCGTAGCGGCAGCACCTGGCAGCTGGGCA 1280
 QY 173 CysSerThrProArgIleHisProPro-----MetGluSerLeuProGlyIleGly 189
 Db 1281 GCCTCCTGCTGGGACATCATCAACCGATTGTCATGCATCGCAAAATACCGCTGGGC 1340
 QY 190 ArgPro-----LysLeuSerAsnAspProAlaGluAspAspLysGluTrpAsnGluLeuTyr 208
 Db 1341 CGCGACTGCGCGATGATGACGAGCTGGA-----CCCTGG-----TAT 1382
 QY 209 SerGluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHis 228
 Db 1383 TCAGAGCTGAAACGAGATTGGCTGCGAGGGCCA-----AAGCATCCCGCCAGACAA 1436
 QY 229 ThrLeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgPro 248
 Db 1437 TCACCCACGAGCGCAGC-----GATATGGTGCCTTTTCCCGCAGCGC 1493
 QY 269 GluAsnLeuPheHisSerIleTyrAsnAspAspLysGlnLysLeu----- 284
 Db 1494 GATAATATTTGCGCAGCGTGTAAACCGCATGGCTATACCTGTGTGGCGCAATAACAAATTGTCTAG 1553
 QY 285 -----PheThrLeuLeuThrAsnHisArgCysThr 294
 Db 1554 GGGCGAAGCAGCTCGCCGTTGGGAAGCAGCCCTACCTGTGTGGCGCAATAACAAATTGTCTAG 1613
 QY 295 ArgLeuAlaLeuThrGlyGlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeu 314
 Db 1614 CCCATTGCGCGATCGGTGCGATGTATAACGCGCATCCATCATGTTGAA----- 1661
 QY 315 LeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMet----- 329
 Db 1662 CGCGCGAAGCAATGGCGGTGTGTGGCGGAGCGGTGTTTATAAAATGGACACC 1721
 QY 329 ----- 329
 Db 1722 GACAGCAATAACCGTATCACCGCGTTCACTGGCTGGATACCTCTGGCGCTTCGCATAA 1781
 QY 330 -----AlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyr 347
 Db 1782 GCCACGGCGAAAGCCTTCGCGCTGCGCTGCAATGCATGCAGACGCGCGCTGTGTGTG 1841
 QY 348 AsnSerGlyPheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsn--- 366
 Db 1842 -----ATGCGCGGAATGACGCTAATCCCAACGGTATCCCGCAACGCT 1883
 QY 367 -----LeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeu 383
 Db 1884 TCCGATATGTTGGCGCGTAACATGATGGACCACTCCGCGCTCCACTGTTCGTTCTGACG 1943
 QY 384 ArgGlnGluPheValAspSerValArgAspProTyrGlyLeuProTrpTrpLysGlu 403

1944 AAAGAA-----CCGCTGTGG----- 1958
 QY 404 AlaValAlaGlnHisIleAlaLysAsnProThr---AspAlaLeuProIleProPheArg 422
 Db 1959 -----CTGGCAAGGTCCGGCGCAACAGCAGCTGTATGTCGTGGCTACCCGT 2003
 QY 423 AspProGlu-ProGlnValThrProPheThrGluGluHisPro----- 437
 Db 2004 GACGGCGACTCCGACGCGCTACTCCGGAACAAGTATCTCAACAATATTTCCTCGT 2063
 QY 438 -TrpHisThrGlnIleHisArg-----AspAlaPheSerTyrGlyAlaValGlyPr 454
 Db 2064 GTGGTCACCGCAACGACACAGCGATGAAAAAGCGCTGGTGGCG- AAAGCACTGGATGA 2122
 QY 454 oGluValAspSerArgValIle-----ValAspLeuArgTrpPheGlyAlaThrAspPr 472
 Db 2123 GGAGATTCGCTACCGCGAGTCCATAGCGTGTATCTCCATCAGTCTGGGAACCGCTGCC 2182
 QY 472 oGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGln 492
 Db 2183 CGATCCGGAACCGCTCTGACGCTGAGCAAAACCGTAAAGATCCGCGATGGCTTGCCTG 2242
 QY 492 nProThrPheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMetAl 512
 Db 2243 CCCCGATATTTATTACG--ACGTGGGTGATTATGTCGTAAGGCGCGGAAGCTTCTCAC 2300
 QY 512 aAspMetCysGluValAlaAlaSerAsnLeuGlyTyrLeuProThrSerProGlnPh 532
 Db 2301 GCGCAGCTGGGACATATCGGCGAGCTGT-----TTGACGCCAAA-GAATT 2344
 QY 532 e---MetAspProGlyLeuAlaLeu-----HisLeuAlaGlyThrThrArgIleGlyPh 549
 Db 2345 CACTATCAGTCAAGCGCTGAATGCCAATAACCAATATGCTGGGTGTTATCATGGGAAA 2404
 QY 549 eAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTy 569
 Db 2405 AAATGCAAAAGAGCGGTGTGTGACGCGCAATTGTGCGGCATTGTGACCAATAATCTCTG 2464
 QY 569 rValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMe 589
 Db 2465 GCTGCGCGGGGAGCGGCATCCCTCCCGCAGCGTGGTGAACAGTACGTACCTGACCATGGC 2524
 QY 589 tCysHisAlaIleLysSerAlaArgSerIleAsnThrLeuLysGly 605
 Db 2525 GCGCTGGGGCTGAAGCGCGCTCATGACATCAGCTGCGCATGAAGGGG 2573

RESULT 11
 AAX57912
 ID AAX57912 standard; DNA; 5187 BP.
 XX
 AC AAX57912;
 XX
 DT 15-JUL-1999 (first entry)
 XX
 DE G. oxydans D-sorbitol dehydrogenase coding sequence.
 XX
 KW D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulonic acid; precursor;
 KW L-ascorbic acid production; ss.
 XX
 OS Gluconobacter oxydans.
 XX
 PN WO920763-A1.
 XX
 PD 29-APR-1999.
 XX
 PF 13-OCT-1998; 98WO-JP04612.
 XX
 PR 17-OCT-1997; 97JP-0285280.
 XX
 PA (FUJI) FUJISAWA PHARM CO LTD.
 XX
 PI Ishii Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;

XX WPI: 1999-302741/25.
 XX Gene group for D-sorbitol dehydrogenase, useful for simple
 PT large-scale production of L-sorbose or 2-keto-L-gulonic acid as
 PT precursor for L-ascorbic acid
 XX Claim 21: Page 60-62; 83pp; Japanese.
 XX This sequence encodes the D-sorbitol dehydrogenase of the
 CC invention. Cells transformed with a vector containing DNA encoding
 CC the dehydrogenase can be used to produce L-sorbose or 2-keto-L-gulonic
 CC acid as precursor for simple large-scale L-ascorbic acid production.
 XX Sequence 5187 BP; 1251 A; 1390 C; 1297 G; 1249 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 1,228-07 Length: 5187
 Score: 206.00 Matches: 149
 Percent Similarity: 30.44% Conservative: 86
 Best Local Similarity: 19.30% Mismatches: 231
 Query Match: 6.27% Indels: 307
 DB: 20 Gaps: 33

US-09-856-327-2 (1-618) x AAX57912 (1-5187)

QY 8 MetLeuArgAspTyrProArgSerMetGlnIleAsnGly----- 20
 DB 1144 ATGATGCGGTCTTCCCAAGACTTATGCGACGAGACCCTTCTACTGGACTGAAAGC 1203
 QY 21 -----GlnIleProLysAsnAlaIleHisGluThrTyr----- 31
 DB 1204 CACCAGTCGTTGAGAGCGCAACAGCGGCCCGGCACTGCTCCATCGGAA-TATGCGCA 1262
 QY 32 -----GlyAsnAspGlyValAsp 37
 DB 1263 GAATCCCAAGTAGAACGGAGTTTATTCATGAGTTCTCGAATTCCTTTCGCGCAGAT 1322
 QY 38 ValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGluAla 57
 DB 1323 GTCGTGATCGTGGATCGGCGTCGACGGGCCAGTATGCGCAACGAACCTTCCGAGAGCC 1382
 QY 58 GlyLeuArgValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsnAla 77
 DB 1383 GGCCTCTCCGTCATCGTTCTTGAAGCCGGCC----- 1415
 QY 78 GluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGluPhe 97
 DB 1416 -----CGGATCGAGCCG 1427
 QY 98 GlnLysAspIleAspArgPhe---ValAsnValIleLysGlyAlaLeuGlnGlnValSer 116
 DB 1428 CAGCATATTCITGAAAAATTCGCGCACCGCAACAGGGGACCATACCGCTTCCCTAC 1487
 QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTrpSerAlaProPro 136
 DB 1488 CCACCCGCG-----CCTGGCGGATGATCGC 1514
 QY 137 GlySerSerAlaIleSerAsnGly-----LysAsnProHisGlnArgGluPhe 152
 DB 1515 CCTGATAGGCGTCTCCCAAGGCTATCTGCATAGCAGCGGACCTCGCGTCTCGGAT 1574
 QY 153 GluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerHisTrpThr 172
 DB 1575 CAG-----CAGGGCTATCTGCGTGTGTCGGGGGACCAACCTGGCATGGGCA 1622
 QY 173 CysSerThrProArgIleHisPro-----ProMetGluSerLeuProGlyIleGly 189
 DB 1623 GGATGTGCTGCGCGGTATCTCCCTCTGACTTCGATTTACATTCGCGATATGGCGTTGCG 1682
 QY 190 ArgProLysLeuSerAsnAspProAlaGluAspLysGlyLutProAsnGluLeuTyrSer 209
 DB 1683 CGC-----GACTGGGCCCATCAAGTACGATGATCTCGAGCCATTCTACTAT 1727

QY 210 GluAlaGluArgLeuIleGlyThrSerThrLysGluPheAspGluSerIleArgHisThr 229
 DB 1728 CAGGCGCAAGTATCATGGCGTGGCAGCCCTAACATGGAT----- 1769
 QY 230 LeuValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePhe-----Arg 247
 DB 1770 -----GTCGATGACCTGGGATCTCCACGATTCACAAATACCCGATGAAG 1814
 QY 248 ProLeuProLeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSer 267
 DB 1815 GAAGTACCCCTGTC----- 1829
 QY 268 AlaGluAsnLeuPheHisSerIleTyrAsnAspLysGlnLysLysLeuPheThrLeu 287
 DB 1830 -----TATGGCGCGGATCAGTTTCGCAACTGATCCATGAG 1865
 QY 288 LeuThrAsnHisArg----- 292
 DB 1866 AAGACGAATTACCGCGTGTTCACGAGCCACAGCGCGTAACACTCGCCCTTATGACAAG 1925
 QY 293 -----CysThrArgLeuAlaLeuThr 299
 DB 1926 CGCCCAACTGTGAGGGCAACAACACTGTCATGCGGATCTGCCGATCGGGCGGATGATC 1985
 QY 300 GlyGly-----TyrGluLysLysIleGlyAlaAlaGluValArgAsnLeu 314
 DB 1986 AACGGAATTCACCTCGGTCAATCATCGGAGCAGCAGCGCCGCTATATTCGATGCG 2045
 QY 315 LeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMet----- 329
 DB 2046 GTTGTCTACCGACTGGAGACCAGCCAGCAGCAAGAAGGTGCTGCCGCCCTAAATATTATC 2105
 QY 330 -----AlaLysValTyrValLeuAlaSerGlyAla 339
 DB 2106 GATCCCGTAGAATTCATCGTGTACCGGTAAGTTCTTCGTGCTGCTGCGCACTGC 2165
 QY 340 Ile-----GlyAsnProGlnIleLeuTyr 347
 DB 2166 ATTGAGAGTGCCAACTGCTCTCTGCTGTCGCGCGATGACAAAATCCCGGGCATGGC 2225
 QY 348 AsnSerGly-----PheSerGlyLeuGlnValThr----- 357
 DB 2226 AACATTCAGATCAGGTTGGTGGCAACATGATGATGATCAGCAGCGGCTACAGCTCTCGTT 2285
 QY 358 ---ProArgAsnAspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnPromet 376
 DB 2286 ATGACCGGAACGACTCTCTGTGGCGGGTCTGCTGCTCTCTGCTGCTGCTGCTGCTGCTG 2333
 QY 377 AlaPheCysGlnIleValLeuArgGlnGluPheValAspSerValArgAspProTyr 396
 DB 2334 -----AGCATTATCGACTCGTTTCGTGACGCG----- 2360
 QY 397 GlyLeuProTrpTrpPheGlu-----AlaValAlaGlnHisIleAlaLysAsnProThr 414
 DB 2361 -----CCATGGCGGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2405
 QY 415 AspAlaLeuProIleProPheArgAspProGluProGlnValThrThrProPheThrGlu 434
 DB 2406 -----GATATCAGGTCGACTTCGCAACGGGCTG----- 2435
 QY 435 GluHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyPro 454
 DB 2436 -----GCGATTGCCAAGGCTATTCGCGGAAA 2462
 QY 455 GluValAspSerArgValIleValAsp-----LeuArgTrpPhe----- 467
 DB 2463 GAGCTGGAAGACGACATCGTTATGGCTCTCTCATGCGGTTCTCTCTTCAGCCATAAC 2522
 QY 468 ---GlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAsp 486
 DB 2523 GAAGCATTTGGCGACCCCGAC-----AACCGGCTGACACTGAGCAAAACACATAAAGAC 2576

Qy	487	GlyTyrSerMetProGlnProThrPheArgTyrArg-----ProSerThrAlaSer	503
Db	2577	GTTCGTGGCATCTCCTCACCCGGAAGCTTATTACAAAGCTTCCGAGTACACAGTGAAGAGT	2636
Qy	504	AsnValArgAlaArgLysMetMetAlaAspMetCysLysValAlaSerAsnLeuGly---	522
Db	2637	TGTGACCATACCAAGAGCTGTTCAGGAAGTCTGATGGCTCTGATGAGTGGTACTGATCCT	2696
Qy	523	-----GlyTyrLeuProThrSerProGlnPheMetAspProGlyLeuAla	538
Db	2697	CAATGGACAAGGGTTACTTCCCG-----CAG	2723
Qy	539	LeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsn	558
Db	2724	TGCCATCCGTGGCGCAGCAGCATATGGGAACAGACCCCAATTCGGTCTGGTACGGT	2783
Qy	559	AsnSerLeuValTrpAspPheAlaAsnLeuTrpValAla-----	571
Db	2784	GAGTGGCGCACCATGACCAAGAAACCTGTTGTTGCCAGATCAGCGGTCTTCTCTTCG	2843
Qy	572	---GlyAsnGlyThrIleArgThrGlyPheGly-----	581
Db	2844	GTCGGTACAGCAATATACCCCTGACCATTGGCGCGCTTCGCGTTCGCGTTCGAGCATCC	2903
Qy	582	-----GlyTyrLeuProThrSerProGlnPheMetAspProGlyLeuAla	587
Db	2904	CTGAAAAAGGAGATGCTTTCATGCGTGAGGGGAATAAAGCGGGAATACGCGCCTCTTCT	2963
Qy	588	SerMetCysHisAla-----IleLysSerAlaArgSerIleIleAsnThrLeu	603
Db	2964	GCCAGCTGCCATAGCTTCGGGTGCTCTGTTTCGGCGCAGTCAGC-----	3008
Qy	604	LysGlyGlyThrAspGlyLysAsnThrGlyGluHis	615
Db	3009	-----GAGGCGAGAGGATCAGGCCACCAC	3032
RESULT 12			
ID	AAx57909		
AA	AAx57909 standard; DNA; 1632 BP.		
AC	AAx57909;		
XX	15-JUL-1999 (first entry)		
XX	G. oxydans D-sorbitol dehydrogenase coding sequence.		
DE	D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulononic acid; precursor;		
KW	L-ascorbic acid production; ss.		
XX	Gluconobacter oxydans.		
XX	WO9920763-A1.		
XX	29-APR-1999.		
XX	13-OCT-1998; 98WO-JP04612.		
XX	17-OCT-1997; 97JP-0285280.		
XX	(FUJI) FUJISAWA PHARM CO LTD.		
PA	Ishii Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;		
PI	WPI; 1999-302741/25.		
DR	P-PSDB; AAY14049.		
XX	Gene group for D-sorbitol dehydrogenase, useful for simple		
PT	large-scale production of L-sorbose or 2-keto-L-gulononic acid as		
PT	precursor for L-ascorbic acid		
XX	Claim 14; Page 52-54; 83pp; Japanese.		
XX	This sequence encodes the D-sorbitol dehydrogenase of the		


```

Db      2520 GGGTACCACCTTCGATACCACCGTGTATCAA-----ACCCACCATAT 2561
QY      541 uAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnSerIle 561
Db      2562 GAGCGCGGGGGCATCATGGGTGAAGATCCGAAACCAGCGAGTGAACCGTATTATGCA 2621
QY      561 uValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheG1 581
Db      2622 GAGCTGGGATGTCGCGAAGCGTGTGTGTCGGGTGCGTCCGGTCCCGAGGGTCTGGG 2681
QY      581 yGluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIle-IleA 601
Db      2682 CTACAACCGGACCGGCATGGTGGCGGCACCTACTGCTGTCGCGCAAGCCATCCGTGA 2741
QY      601 snThrLeu-----LysGlyGlyThrAspGlyL 610
Db      2742 ACAGTATCTGAAGAACCCAGGTCCACTGTCGAGGCATGAAGAAACGGCATGATGAA 2801
QY      610 ysAsnThrGly 613
Db      2802 AGCATTCGGC 2812

```

RESULT 14

ABL52744

ID ABL52744 standard; DNA; 1689 bp.

XX ABL52744;

AC

XX

DT 01-JUL-2002 (first entry)

XX Glucose-3-dehydrogenase encoding sequence.

DE

XX Glucose-3-dehydrogenase; G3DH; enzyme; ds.

KW

XX

OS Halomonas sp. alpha-15.

XX

FH Key

FT CDS

FT

FT

XX

PN JP2002017372-A.

XX

PD

XX

PF 30-JUN-2000; 2000JP-0237709.

XX

PR 30-JUN-2000; 2000JP-0237709.

XX

PA (HAYA/) HAYADE K.

XX

DR WPI; 2002-221711/28.

XX

DR P-PSDB; ABB09452.

XX

PT Glucose-3-dehydrogenase and its preparation

XX

PS Claim 5; Page 9-10; 16pp; Japanese.

XX

CC The invention relates to Glucose-3-dehydrogenase (G3DH) enzyme. The

CC methods of the invention are used for the preparation of G3DH. The

CC current sequence represents the glucose-3-dehydrogenase encoding

CC sequence.

XX

SQ Sequence 1689 BP; 380 A; 446 C; 508 G; 355 T; 0 other;

XX

```

US-09-856-327-2 (1-618) x ABL52744 (1-1689)
QY      37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
Db      19 GATGCCATCGGTGGTGGCTCAGGCATTAAGCGCGGTGGTGGCTGCTAAAGAGTAACACCG 78
QY      57 AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
Db      79 AAGGGCTAAAGGTTTCTGCTT----- 102
QY      77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLys----- 91
Db      103 ---GAGCGTGGCGCGCAATATTAGCACGTAAAGGATTACCAATGCGGCAAGAGGCG 159
QY      92 -----LysAsnGluIleGluPheGlnLysAspIleAsnArgPheValAsn 106
Db      150 TGGGATTACCCACCACCGTAATGAG---CCAACCCAGGAATGATCGCTAATAT----- 210
QY      107 ValIleLysGlyAlaLeuGlnValSerValProValArgAsnGlnAsnValPro--- 125
Db      211 -----CCTGTGCTCAAGCGCGACTACCCCTT 237
QY      126 -----ThrLeuAspProGlyAlaTrpSerAlaProProGlySerSerAlaIleSer 142
Db      238 AATGAAGCCACCCCTG---GGCATGTGGCG-----GAT 267
QY      143 AsnGlyLysAsnProHis-----GlnArgGluPheGluAsnLeuSerAlaGluAlaVal 160
Db      268 GAACAGGCCAATCCCTACGTAGAGAGAAAGCGCTTCGACTGGTTT----- 312
QY      161 ThrArgGly-----ValGlyGlyMetSerThrHisTrpThrCysSerThrProArgIle 178
Db      313 ---CGCGCTACCGTAGCGCGCTCTCTGCTTGGGGTGGCGCAAGCTACCGGCTC 369
QY      179 HisProMetGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAsp----- 196
Db      370 AGCCCG-----ATGGATTTTGAGCCCAACAGCGGGAAGGCGCATCGCCATTGATGGCG 423
QY      197 ---ProAlaGluAspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeuIle 215
Db      424 ATTCGCTACGAAGATCTCGCCCGCTGG-----TATGACTATGTGGAGCGCTTGGC 474
QY      216 Gly---ThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValLeuArgSer 234
Db      475 GGTATCGCAGCAGCAGCAGGAG-----GGCTGGATATC 507
QY      235 LeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCys--- 253
Db      508 CTTCGGATGGC-----GAGTTTCTGCGCGCAATACCGCTTAACCTGTGTG 552
QY      254 -----HisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerAla 268
Db      553 GAAGAGGATGCGGCCAAGCGTATTAAAGAGGCG-----TTTGGTGGCGAG 597
QY      269 GluAsnLeuPheHisSer-----IleTyrAsnAspLysGlnLysLysLeu 284
Db      598 CGTCACTTATCCACAGCGGGTGGCCCAATATTACCCAGCCCAAGAGCAACACCGC 657
QY      285 PheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGlyGlyTyrGluLys 304
Db      658 GTCAATTGCCAATACCGAAATAAATGCTGGCTGGGTGCTCCCTTATGTGGTGGCTAT----- 711
QY      305 LysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThrArgAsn-----ProSer 321
Db      712 -----TTACAGACCCAGTCCGCTACGCTTCCCGCG 741
QY      322 SerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGly 341
Db      742 CGGTGGCCACAGGTAATCTACGTACGGCCATCTCGATTCTGCAGC----- 789
QY      342 AsnProGlnIleLeuTyrAsnSerGly-PheSerGlyLeuGlnValThrProArgAsnAs 361

```

```

Alignment Scores:
Pred. No.: 1.69e-05 Length: 1689
Score: 174.00 Matches: 141
Percent Similarity: 33.18% Conservatives: 79
Best Local Similarity: 21.27% Mismatches: 242
Query Match: 5.30% Indels: 201
DB: 24 Gaps: 33

```

Db 790 -----CAGTGTGTACGACAAAGACCGTCTACGGCGCGGGCGGTG----- 832
Qy 361 pSerLeuIleProAsnLeuGlyArgTyrIleThrGluInProMetAlaPheCysGlnI 381
Db 833 -----AGTGATTGAGCTGAGACCCATGAAGTCCATGAGTAC 870
Qy 381 eValLeuArgGlnGluPheValAspSerValArgAspProTyrGlyLeu----- 398
Db 871 ACCGTGTGTCTCTCTCAATCGTCTACGTTCAACACACCTGGATTGATGAAC 930
Qy 399 -----ProTrp-----TrpLysGluAlaValAlaGlnHisIleAlaL 411
Db 931 TCGGCCACCGATGCTGGGAAGTGGCTGGGACGACGAGTAGTGGCAACTGGGCCACA-- 988
Qy 411 sAsnProThr-AspAlaLeuProIleProPheArg-----A 423
Db 989 -----ACGTGATGATCACCACCTCCGCTGCGGTGCCAGTGGCGAAGTGAAGCTAT 1041
Qy 423 spProGluProGlnValThrProPheThrGlu---GluHis-ProTrpHisThrGln 441
Db 1042 CTCGACAAATACTACTCTCGACGCCGCTCTCGGGCTTTTACATTCCGCGCTTTGCGCAAC 1101
Qy 442 IleHisArgAspAlaPheSerTyr-----GlyAlaValGlyPro 454
Db 1102 GTAGCGGATGAGCAGCAGCAAGCTATGCGCGGTTCGGTTATCAAGCGCGCGACGCCG 1161
Qy 455 GluValAspSerArgValIleValAspLeuArg----- 465
Db 1162 GAGGGCTGGATCGCGAAATGCTGAGCTCAATATCGCGCGCTGATTAAAGCAAGCACATC 1221
Qy 466 -----Trp-----PheGlyAlaThrAspProGluAla 474
Db 1222 ACCACGCCAGCGGTGGACGATCGGTATGACGGGCTTCGGCGAAATGCTGCCGATCAC 1281
Qy 475 AsnAsnLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThr 494
Db 1282 GATAACCGTATTCTCTCGATCATAGCTTCGCGATAAATGGGGCTCGCGTGTGTCG 1341
Qy 495 PheArgTyrArgProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMet 514
Db 1342 ATCGACGTTGAGTCAAGCAAAAGCGCGGATATCGTCCGCGACATGTTCAAGATGG 1401
Qy 515 CysGlu-----ValAlaSerAsnLeuGlyGlyTyrLeu 525
Db 1402 GTCGATCTCTAGAGCGCTGCGGTGAAGAAGCTCAAAAGCGACGTAGCGGATTCACGCA 1461
Qy 526 ProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThr 545
Db 1462 -----CCGGCATGGGATCCATCAAAATGGGCGACTGCA 1494
Qy 546 ArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPhe 565
Db 1495 CGGATGGCGGTGACCCGAAACGTCAGTACTCAATAGCCATACACAGGTGTGGACGG 1554
Qy 566 AlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThr 585
Db 1555 CCCAATGTGTTGTCACCGATGGCGCTGCATGACGCTCATCTCTTTCGCGTAACCTTCT 1614
Qy 586 LeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsnThrLeuLysGly 605
Db 1615 CTAACCTATATGGCGTTACCGCGCGGGCGGTGGATTATGCGCGTGGAGAGCTGAAGCGG 1674
Qy 606 Gly 606
Db 1675 GGG 1677
RESULT 15
AAC36995
ID AAC36995 standard; DNA; 2484 BP.
XX
AC AAC36995;
XX
DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 15807.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-013256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.

Query Match

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 01:52:53 ; Search time 4438 Seconds

(without alignments)

11533.399 Million cell updates/sec

Title: US-09-856-327-1

Perfect score: 2106

Sequence: 1 atcagccatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_nam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gsl.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97.4	4.6	630	14	CA349327 619706 NC
2	96.2	4.6	1201	13	BX446296 BX446296
3	95.8	4.5	553	14	CB054499 NISC_gm04
4	95.4	4.5	336	12	BM884782 rcl15a03.y

C	5	95.4	4.5	480	9	AI392214	AI392214 NCSPIF9T7
C	6	95.2	4.5	871	9	AL514867	AL514867 AL514867
	7	95	4.5	271	13	BU531092	BU531092 AGENCOURT
	8	95	4.5	503	11	AY109012	AY109012 Zea mays
	9	94.8	4.5	509	13	BQ276784	BQ276784 AGENCOURT
	10	94.2	4.5	165	12	BI746603	BI746603 rm29d01.y
	11	94.2	4.5	286	10	BG736597	BG736597 rk59f08.y
	12	94	4.5	974	13	BX327262	BX327262 BX327262
	13	93.8	4.5	318	12	BI941623	BI941623 sc80a12.y
	14	93.8	4.5	381	13	BX425197	BX425197 BX425197
C	15	93.8	4.5	417	13	BQ526732	BQ526732 NISC_no18
C	16	93.6	4.4	263	12	BM900884	BM900884 rc42f02.y
	17	93.6	4.4	809	14	CA453483	CA453483 AGENCOURT
	18	93.6	4.4	911	13	BX380015	BX380015 BX380015
C	19	93.6	4.4	1104	9	AL513809	AL513809 AL513809
	20	93.4	4.4	162	9	AW395255	AW395255 SH45g11.y
C	21	93.4	4.4	1201	9	AL513901	AL513901 AL513901
	22	93.2	4.4	436	14	CA799440	CA799440 sat33g04.
	23	93.2	4.4	484	12	BI680498	BI680498 458776 MA
	24	93.2	4.4	949	13	BO433605	BO433605 AGENCOURT
	25	93.2	4.4	1017	14	CB946057	CB946057 AGENCOURT
	26	93	4.4	248	13	BU589063	BU589063 AGENCOURT
	27	93	4.4	600	12	BG926622	BG926622 HNC56-1-H
	28	93	4.4	685	9	AV714975	AV714975 AV714975
	29	93	4.4	936	13	BU850422	BU850422 AGENCOURT
	30	93	4.4	1334	12	BM555708	BM555708 AGENCOURT
	31	92.8	4.4	299	14	CA688814	CA688814 vlm96.pk0
	32	92.8	4.4	344	12	BM514150	BM514150 kx96e09.y
	33	92.8	4.4	593	12	BM514170	BM514170 kx97a03.y
	34	92.6	4.4	262	10	BE722980	BE722980 192270 MA
	35	92.6	4.4	314	14	CA802494	CA802494 sau37b11.
	36	92.6	4.4	337	12	BI142697	BI142697 kt44f06.y
C	37	92.6	4.4	409	13	BX425137	BX425137 BX425137
	38	92.6	4.4	421	10	BG362036	BG362036 gb50f09.y
	39	92.6	4.4	586	14	CA370380	CA370380 650538 NC
	40	92.6	4.4	778	13	BU533506	BU533506 AGENCOURT
	41	92.6	4.4	347	13	BU834850	BU834850 T066E04 P
	42	92.4	4.4	434	14	CA849260	CA849260 k109h12.y
	43	92.4	4.4	620	14	CA353533	CA353533 625052 NC
	44	92.4	4.4	844	13	BU908877	BU908877 AGENCOURT
	45	92.2	4.4	308	10	BE633260	BE633260 uv74g09.y

ALIGNMENTS

RESULT 1

CA349327

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA349327 619706 NCCWA LRT Oncorhynchus mykiss cdna clone lrt140m18_C_G09
5', mRNA sequence.

CA349327.1 GI:24594498

EST.

Oncorhynchus mykiss (rainbow trout)

Eukaryota; Metazoa;

Actinopterygii; Neopterygii;

Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 630)

Rexroad,C.B. and Keeler,J.W.

Sequence analysis of a rainbow trout normalized cdna library

Unpublished

Contact: Rexroad CE

USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccwa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim_alt option. Vector identified by

cross_match v0.990329.

Seq primer: AGCGGATAACAATTTTCACACAGGA.

PR	21-JUN-1999;	9905-011398617
PR	22-JUN-1999;	9905-0133859
PR	23-JUN-1999;	9905-0140353
PR	24-JUN-1999;	9905-0140354
PR	28-JUN-1999;	9905-0140695
PR	29-JUN-1999;	9905-0140951
PR	30-JUN-1999;	9905-0141287
PR	01-JUL-1999;	9905-0141842
PR	01-JUL-1999;	9905-0141854
PR	02-JUL-1999;	9905-0142305
PR	06-JUL-1999;	9905-0143750
PR	08-JUL-1999;	9905-0142803
PR	09-JUL-1999;	9905-0142920
PR	12-JUL-1999;	9905-0142977
PR	13-JUL-1999;	9905-0143524
PR	14-JUL-1999;	9905-0143642
PR	15-JUL-1999;	9905-0144005
PR	16-JUL-1999;	9905-0144085
PR	16-JUL-1999;	9905-0144086
PR	19-JUL-1999;	9905-0144335
PR	19-JUL-1999;	9905-0144331
PR	19-JUL-1999;	9905-0144332
PR	19-JUL-1999;	9905-0144333
PR	19-JUL-1999;	9905-0144334
PR	19-JUL-1999;	9905-0144335
PR	20-JUL-1999;	9905-0144352
PR	20-JUL-1999;	9905-0144632
PR	20-JUL-1999;	9905-0144684
PR	21-JUL-1999;	9905-0144814
PR	21-JUL-1999;	9905-0145086
PR	21-JUL-1999;	9905-0145087
PR	22-JUL-1999;	9905-0145085
PR	22-JUL-1999;	9905-0145087
PR	22-JUL-1999;	9905-0145089
PR	22-JUL-1999;	9905-0145119
PR	23-JUL-1999;	9905-0145145
PR	23-JUL-1999;	9905-0145218
PR	23-JUL-1999;	9905-0145224
PR	26-JUL-1999;	9905-0145276
PR	27-JUL-1999;	9905-0145913
PR	27-JUL-1999;	9905-0145918
PR	27-JUL-1999;	9905-0145919
PR	28-JUL-1999;	9905-0145951
PR	02-AUG-1999;	9905-0146386
PR	02-AUG-1999;	9905-0146388
PR	02-AUG-1999;	9905-0146389
PR	03-AUG-1999;	9905-0147038
PR	04-AUG-1999;	9905-0147204
PR	04-AUG-1999;	9905-0147302
PR	05-AUG-1999;	9905-0147192
PR	06-AUG-1999;	9905-0147260
PR	06-AUG-1999;	9905-0147303
PR	06-AUG-1999;	9905-0147416
PR	09-AUG-1999;	9905-0147493
PR	09-AUG-1999;	9905-0147935
PR	10-AUG-1999;	9905-0148171
PR	11-AUG-1999;	9905-0148319
PR	12-AUG-1999;	9905-0148341
PR	13-AUG-1999;	9905-0148654
PR	13-AUG-1999;	9905-0148684
PR	16-AUG-1999;	9905-0148368
PR	17-AUG-1999;	9905-0149175
PR	18-AUG-1999;	9905-0149426
PR	20-AUG-1999;	9905-0149722
PR	20-AUG-1999;	9905-0149723
PR	20-AUG-1999;	9905-0149929
PR	23-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150884
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066

PR	27-AUG-1999;	9905-0151080
PR	30-AUG-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151438
PR	01-SEP-1999;	9905-0151900
PR	07-SEP-1999;	9905-0151363
PR	10-SEP-1999;	9905-0153070
PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039
PR	20-SEP-1999;	9905-0154779
PR	22-SEP-1999;	9905-0155133
PR	23-SEP-1999;	9905-0155486
PR	24-SEP-1999;	9905-0155659
PR	28-SEP-1999;	9905-0155458
PR	29-SEP-1999;	9905-0155596
PR	04-OCT-1999;	9905-0157117
PR	05-OCT-1999;	9905-0157753
PR	06-OCT-1999;	9905-0157865
PR	07-OCT-1999;	9905-0158029
PR	08-OCT-1999;	9905-0158232
PR	12-OCT-1999;	9905-0158393
PR	13-OCT-1999;	9905-0155293
PR	13-OCT-1999;	9905-0158294
PR	14-OCT-1999;	9905-0155329
PR	14-OCT-1999;	9905-0155329
PR	14-OCT-1999;	9905-0155331
PR	14-OCT-1999;	9905-0155638
PR	18-OCT-1999;	9905-0155638
PR	18-OCT-1999;	9905-0155584
PR	21-OCT-1999;	9905-0160741
PR	21-OCT-1999;	9905-0160767
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160770
PR	21-OCT-1999;	9905-0160814
PR	21-OCT-1999;	9905-0160815
PR	22-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160989
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161406
PR	26-OCT-1999;	9905-0161359
PR	26-OCT-1999;	9905-0161360
PR	26-OCT-1999;	9905-0161361
PR	28-OCT-1999;	9905-0161992
PR	28-OCT-1999;	9905-0161993
PR	29-OCT-1999;	9905-0162142

Alignment Scores:	
Pred. No.:	0.000548
Score:	19.50
Percent Similarity:	33.58%
Best Local Similarity:	19.23%
Query Match:	4.86%
DB:	21
Gaps:	30
Length:	2484
Matches:	130
Conservative:	97
Mismatches:	224
Indels:	225

[illegible]

```

Db      902 -----AACCTACTTTACACCTCATGACTATTCGGGCTTACAGAAAGTTCCGCTATGCTCGAG 955
Qy      89  TYNHLSLYSASNGLIIEGLUPHEGLNLYSAPRIIEASPRGHEVALSNVALIE 108
Db      956  TTTATCGAAAAAGC----- 970
Qy      109  LYSGLIYALALEUGLNGINLSERVALPROVALARGASNGINASNVALPROTHREUASP 128
Db      971  ---GGCCTTTGTGCTACGGTT----- 988
Qy      129  PROGLIYALATPSSERIALAPROPROGLYSERSERIALIIESERASNGIYLYSAPROHIS 148
Db      989  -----GATGGGAAA----- 997
Qy      149  GINARGLUPHEGLIUBASNLEUSERIALAGLUALAVALTHRARGLYVALIGLYMETSER 168
Db      998  -----TTTATGCTCTGGCTGGTCACTGTGGG-----GGAGGTACAGCT 1039
Qy      169  TNRHISTPRPHCYSETHRPRARGILEHISPROIROMETGLUSERLEUPROGLIYLE 188
Db      1040  GTTAATGCTGTCATCT-----ATPAGGACACCGCATCATGTTTG----- 1081
Qy      189  GLIYARGPROLYSLEUSERASNSPROMIAGLUIASPSPLYSGLUTRPSNGLIUEUTYR 208
Db      1082  -----CAGCAATG----- 1090
Qy      209  SERGLIUALAGLUIARGLEULIEGLYTHRSETHRTHLSGLUPHE-----ASPLU 224
Db      1091  TCTGAAGGAGTAAAGATCAAGTTTGTGTACCAAGATATCACAGTCTCATGTGATGAA 1150
Qy      225  ---SERIIEARGHISTHREUVALLEUARGSERLEUGLNASPALATYRILYSAPRGLN 243
Db      1151  GTTAGCATGAAGATTGCTGTACGGAAGATGTCAAAACACGGGTTTCAGAT----- 1204
Qy      244  ARGILEPHEARGPROLEUPROLEUALACYSHISARGLEU----- 256
Db      1205  -----CAGGTTCTCGGAAAGGCTCGAGAGCTCGTTTCAGAGGTACAAATCAGTT 1255
Qy      257  ---LYSASNALAPROGLU-----TYRVALGLUTRPHISER 267
Db      1256  CCAAGGAACCTACACAGAGATCATCTGTGCTGTGGGTATGAGATGAGAGAGAGA 1315
Qy      268  ALAGLUIASNLEUPHEHISERILETYRASNASPAPLYSGINLYSLEUPHEHREU 287
Db      1316  GCCAAGACGCGAGATCAAACTGCTGTGTGATGCTGTAGAGATGTCACAGTATC 1375
Qy      288  LEUTHRASNHISARGCYSTRHARGLEUALALEU-----THRGLYGLYTRGLU 303
Db      1376  CTGACAGGGATCAAAAGCTGAGAGATTGTATTAGTAGACACACTRAGTAGCAATGAG 1435
Qy      304  LYSLYS-----IIEGLIYALAGLUIVALARGASNLEUENUALATHIRARG 318
Db      1436  AGAAGAAAGCATGTGTGGAGATCTTGTCTACTCTGTGGAGAAAGTATGGAAAGAG 1495
Qy      319  ASNPROSERSEGLIUEASPSERTYRILEMETALALYSVALTYRVALLEUALASERGLY 338
Db      1496  -----TTTATTATCGAAGCTCGAGTAAACGTTTCATCCCTGGA 1534
Qy      339  ALAIIIEGLYASNPROGLINILEUUTYRASNSERGLYPHESERGLYLEUGLINVALTHRPRO 358
Db      1535  TCGCTGTAAACACCGCTCTGTGATGCTTTCGAGTGGGCTAAAGAAC----- 1579
Qy      359  ARGASNASPSERLEUIEPROASNLEUGLIYARG----- 369
Db      1580  -----CCGAACATGGAAGAACTTAAAGCTGCACCCGGTCTGTGATG 1621
Qy      370  -----TYRILETHRCGLINIPROMETALAPHECYSGLINILEVALLEUARGINLU 386
Db      1622  ACATGGGATACTTCCCGAGAAAGATTTCTGATTTCTCCGGAATAATGATGAGGAGGG 1681
Qy      387  PHEVALASPSERVALARG-----ASAPSPROTLYRGLYLEUPROTTRIPLYSGIUALA 404

```

```

Db      1682  ATCATCAGCTCCGTCACATATGATGAT-----ACTGAATCT 1720
Qy      405  VALIAGLINHISILEALALYSASNPROTHRASPALALEUPROILEPROHEARGASPRO 424
Db      1721  GGATCGAAAGCAATTCGTGAGATTCACACGATAGAGA---CGACGTTCTTACGACAGGTTG 1777
Qy      425  GIUPROGLINVALTHRTHRPHOETHRGLIUNHISPROTPRPHISTHRCINLEHISARG 444
Db      1778  AGCCCGTGGGTTTCA----- 1792
Qy      445  ASPALAPHESERTYGLIYALAVAGLYPROGLIUALASPSERARGVALILEVALASPLEU 464
Db      1793  -----GGACCGACCTTAAAGAGAGGATGATTAG----- 1822
Qy      465  ARGTRPHEGLIYALATHRASPRPROGLIUALAASNLEUENUALPHEGLINASNVAL 484
Db      1823  -----TAGGAAAGAAC-----GCTCATCTTTTGTGCTGTGGGACTTA 1864
Qy      485  GLIASPGLYTYRSEMET-----PROGLINPROTHRPHIARGTYRARGPROSERTHRALA 502
Db      1865  GGGTCAGGTGAGGTTATGATGGAACCAAGTTACATACAGGACAACCAAGAAAGATAGA 1924
Qy      503  SERASNVALARGALARGYSMETMETALASPMETCYGLIYVALIASER----- 519
Db      1925  GAGAAITCTGAGGGCAGGCTTACACAAAGCTTTGCGGGTTTGGTTGCAAGCTGTGCAGTT 1984
Qy      520  ASNLEUGLYLYTYR-----LEUPROTHR 527
Db      1985  GAAGTGCGAACAATNAGAGATGATGACAGAAAGATGAAGTNGAGGCAATTACTAAAGA 2044
Qy      528  SERPROPROGLINPHIEMETASPROGLIUEUALALEU----- 539
Db      2045  GCCATGAAAGATTTCTAGACGAAGTTGATGCTGTGTGGGTTGGCACAAGAGAGAG 2104
Qy      540  -----HISLEUALAGLYTHRTHRARGILEGLYHEASPLY 551
Db      2105  TATTGGACGACGTACTTCTCGGCTCACGATGGGTAGTTGAGATGGAGATCACCCG 2164
Qy      552  ALATHRTHRVALALASPSASNSENLEUVALTRPASPHEALASNLEUTYRVALA 571
Db      2165  GAGAAAGTGCCTGTGATGAGATGAGTGTGAGACCTGGGAAGCAGAGGGGTTGTGTTGT 2224
Qy      572  GLIYASNGLYTHRIIEARGTHRCGLYPHEGLIUIASNPROTHRLEUTHSERMETCYSHIS 591
Db      2225  GATGGAAGTATATTCATACAGCTGTGGGTTTAAACCAATGATCACCATTCAGTCTACT 2284
Qy      592  ALAIIIEYSERVALARGSERILEIASNTHREULYSGLIYLYTHR 607
Db      2285  GCTTACTGATCTCCTCAAAAGATTGACTGCTGCAAAACAAACA 2332

```

Search completed: September 24, 2003, 06:06:14
 Job time : 558 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: September 24, 2003, 05:32:15 ; Search time 6374 Seconds
(without alignments)
3966.450 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSFGMRDYPKSMQING.....IINTLKGTDGKNTGHEHRL 618

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh
-Q=/cgn2_1/USFTO.spool/US09856327/runat.23092003_153419_15506/app_query.fasta.1.775
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09856327 @cgn2_1.13963 @runat.23092003_153419_15506 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WANT_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg.hum:*
31: em.htg.in:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rtd:*
36: em.htg.mam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3284	100.0	2106	6 BD013176	BD013176 A novel p
2	1239.5	37.7	2064	8 AF535193	AF535193 Penicophor
3	1239.5	37.7	2069	8 AY291124	AY291124 Tramefes
4	1232.5	37.5	1869	6 E11766	E11766 CDNA of pyr
5	1232.5	37.5	1869	6 I82410	I82410 Sequence 1
6	1232.5	37.5	2046	8 D73369	D73369 Cortolus ve
7	1232	37.5	1902	6 E32737	E32737 Panus cornu
8	1210	36.8	1869	6 AR141573	AR141573 Sequence
9	1091	33.2	1701	6 AR170485	AR170485 Sequence
10	1044	31.8	1946	6 E33665	E33665 Novel prote
11	1044	31.8	1946	6 E37379	E37379 Reagent for
12	1044	31.8	1946	6 AB043883	AB043883 Trichobac
13	243	7.4	165137	1 ANI507836	ANI507836 Arthrobac
14	227	6.3	305520	1 AE016780	AE016780 Pseudomon
15	208.5	6.3	2465	1 AF430844	AF430844 Burkholde
16	208.5	6.3	2467	6 BD168939	BD168939 Novel glu
17	207.5	6.3	299700	1 AP005944	AP005944 Bradythiz
18	207	6.3	4070	1 AF068066	AF068066 Erythrina h
19	206	6.3	198050	1 AB039821	AB039821 Glucosoba
20	205.5	6.3	198050	1 AL646061	AL646061 Ralstonia
21	200.5	6.1	308601	1 CJ11168X2	CJ11168X2 Campyloba
22	198	6.0	4665	1 EC97665	EC97665 Erythrina cyp
23	198	6.0	4665	6 AR182220	AR182220 Sequence
24	191.5	5.8	1845	6 AR182218	AR182218 Sequence
25	190.5	5.8	13036	1 AE005839	AE005839 Caulobact
26	190	5.8	349050	1 AP003586	AP003586 Nostoc sp
27	186	5.7	13508	1 AE005037	AE005037 Halobacte
28	183.5	5.6	11558	1 AE004652	AE004652 Pseudomon
29	180	5.5	4349	1 AF131202	AF131202 Pantoea c
30	177	5.4	299800	1 AP005028	AP005028 Streptomy
31	176.5	5.4	213732	1 AE001862	AE001862 Deinococc
32	174	5.3	1689	6 BD107473	BD107473 Glucose-3
33	170	5.2	11332	1 AE011848	AE011848 Xanthomon
34	162	4.9	301214	1 AE016786	AE016786 Pseudomon
35	160	4.9	339681	1 AP003009	AP003009 Mesorhizo
36	159.5	4.9	2484	8 AY085885	AY085885 Arabidops
37	158.5	4.8	2247	8 ATH316231	ATH316231 Arabidops
38	158.5	4.8	2278	8 BT001942	BT001942 Arabidops
39	158.5	4.8	2385	8 AY039977	AY039977 Arabidops
40	158.5	4.8	43277	2 T5F17	T5F17 Arabidops
41	158.5	4.8	197655	8 ATCHRI69	ATCHRI69 Arabidops
42	157	4.8	10063	1 AE009366	AE009366 Agrobacte
43	157	4.8	10484	1 AE008247	AE008247 Agrobacte
44	152.5	4.6	13405	1 AE014492	AE014492 Brucella
45	149	4.5	332635	1 AP003005	AP003005 Mesorhizo

RESULT 1

ALIGNMENTS

BD013176 2106 bp DNA linear PAT 02-AUG-2002
LOCUS
DEFINITION A novel protein, a gene coding therefor and a method of using the same.
ACCESSION BD013176
VERSION BD013176.1 GI:22093365
KEYWORDS WO 0121657-A/1.
SOURCE Lyophyllum shimeji
ORGANISM Lyophyllum shimeji
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Tricholomataceae; Lyophyllum.
REFERENCE 1 (bases 1 to 2106)
AUTHORS Takakura,Y., Kuwata,S. and Inoue,Y.
TITLE A novel protein, a gene coding therefor and a method of using the patent: WO 0121657-A 1 29-MAR-2001;
JOURNAL JAPAN TOBACCO INC., CORPORATE JURIDICAL PERSON SOCIETY FOR TECHNO INNOVATION OF AGRICULTURE YUKIO SHIMOMURA, SHIRO TAKEKAWA, NOBUO CHO FORESTRY AND FISHERIES, YOSHIMITSU TAKAKURA, SHIGERU KUWATA, YASUHIRO INOUE
COMMENT OS Lyophyllum shimeji
PN WO 0121657-A/1
PD 29-MAR-2001
PE 20-SEP-2000 WO 2000JP006404
PR 21-SEP-1999 JP 99P 267238
PI YOSHIMITSU TAKAKURA, SHIGERU KUWATA, YASUHIRO INOUE PC
CO7A14/375,C12N15/31,C12N15/63,C12N1/21,C12Q1/58,C12P21/02, PC
A01M65/00
CC
FH Key Location/Qualifiers
FT CDS (8)..(1881).
FEATURES
source 1..2106
Location/Qualifiers
1..2106 /organism="Lyophyllum shimeji"
/mol_type="genomic DNA"
/db_xref="taxon:47721"
BASE COUNT 593 a 540 c 527 g 446 t
ORIGIN
Alignment Scores:
Pred. No.: 3,31e-247 Length: 2106
Score: 3284.00 Matches: 618
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-09-856-327-2 (1-618) x BD013176 (1-2106)
QY 1 MetSerLeuSerThrGluGlnMetLeuArgAspTyrProArgSerMetGlnIleAsnGly 20
DB 8 ATGCTCTCTCAACCGAGAGATGCTACCGACTATCCACGCTCATGCAAAATCAACGGA 67
QY 21 GlnIleProLysAsnAlaIleHisGluThrTyrGlyAsnAspGlyValAspAlaPheIle 40
DB 68 CAGATTCTTCAAGACGCAATTACGAAACATACGAAACGACGAGGAGTTGATTTATTCATT 127
QY 41 AlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuGlyValGlnAlaGlyLeuArg 60
DB 128 GCAGATCTGCAGCCATTCGACGACGATGCAAAAGCTCTGTGGAAGCTGCTACGTT 187
QY 61 ValValMetValGlnIleGlyAlaAlaAspSerPheTyrAlaValAlaAsnAlaGluGly 80
DB 188 GTTGTGATGCTGAGATCGGAGCTCTGTATGCTTCTAGCTGTAAATGCCAAGAGAGA 247
QY 81 ThrAlaValProTyrValProGlyTyrHisLysLysAsnGlnIleGluPheGlnLysAsp 100
DB 248 ACTGCAAGTTCCTTACGTTCTCTGCTACCAACAAGAAATGAAATGAGATTCACGAAGAAT 307
QY 101 IleAspArgPheValAsnValIleLysGlyAlaLeuGlnValSerValProValArg 120
DB 308 ATTGACGCGTTCGCAATGATCAAGGAGCGCTTACCAACAAGTCTCTGTTCTGTCAGA 367
QY 121 AsnGlnAsnValProThrLeuAspProGlyAlaTyrPheAlaProGlySerSerAla 140

DB 368 AACCAAGACGCTGACACTTATCCCGGAGCGCTGGAGCGCCCGCTGGAGTTCCAGCC 427
QY 141 IleSerAsnGlyLysAsnProHisGlnArgGluPheGlnAsnLeuSerAlaGluAlaVal 160
DB 428 ATATCGAAGCGTAAAAATCTCTACCAACGCGGAATTCGAGAACTTGTCTCCGAGCCGTA 487
QY 161 ThrArgGlyValGlyGlyMetSerThrHisTyrPheCysSerThrProArgIleHisPro 180
DB 488 ACGCGTGGATGGCGGCGATGAGTACCCACTGAGCTGTCTCCACCGACGAGTATCATCA 547
QY 181 PrometGluSerLeuProGlyIleGlyAlaGlyProLysLeuSerAsnAspProAlaGluAsp 200
DB 548 CCCATGGAAGTCTCCCGGCGATCGCGCTCCGAAGCTCAGTACGACCGCGCAAGAGAC 607
QY 201 AspLysGluTyrAsnGlnLeuTyrSerGluAlaGluArgLeuIleGlyThrSerThrLys 220
DB 608 GACAAAGAGTGGAAACGAGCTTATTCGAGGCGGAGCGCTCTATGCGGACTTCCACCAAG 667
QY 221 GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys 240
DB 668 GAATTCGACGAGTCAATTCGGCACACCTTGTTCGGCTTTTGCAAGACGCTACCAAG 727
QY 241 AspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLeuLysAsnAlaPro 260
DB 728 GATCGTCAACGTAATCTTTCGCGCTCCGCTCCGCTGCGATGCGACCGGTTGAAGACCGCG 787
QY 261 GluTyrValGluTyrHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLys 280
DB 788 GAATTCGTCGATGGCAGCTACAGCAAAATCTTTCCACTCTATCTACACGATGACAAG 847
QY 281 GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly 300
DB 848 CAGAAGAAGCTTATACCTGCTGACGAACCATCTGCTGACACGACTGGCGCTTACGGGC 907
QY 301 GlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuAlaThrArgAsnPro 320
DB 908 GGGTATGAGAAAGATATGGCGCTCGCGAGGTACAGAACTTACTGGCCACGAAATCT 967
QY 321 SerSerGlnLeuAspSerTyrIleMetAlaLysValTyrValLeuAlaSerGlyAlaIle 340
DB 968 AGTTGGCAGCTGACAGCATATATCATGCGGAAGATATATGTAAGTGGCGCGGAGCGATC 1027
QY 341 GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn 360
DB 1028 GGCAACCCACAGATTTCTATTAATCGGGCTCTCGGCGTACAGTCCAGCCACGCAAT 1087
QY 361 AspSerLeuIleProAsnLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln 380
DB 1088 GACTCGTTGATCCCAACCTGGGAGGTACATCAGGAGCAGCAGCCGATGTCATTTGGCAG 1147
QY 381 IleValIleArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProTrp 400
DB 1148 ATAGCTTGAGGACAGAAATTCGTCACAGCGCGCGCGAGATCCTTATGAGACTGCCATG 1207
QY 401 TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro 420
DB 1208 TGGAAAGAACCGCTGCTCAACATATGCGCAAGAACCCGACAGATGACCTGCCATTCGG 1267
QY 421 PheArgAspProGluProGlnValThrThrProPheThrGluGlnHisProThrPheHis 440
DB 1268 TTCGCGCATCGGAACCCCGAGGTAAACACCCCATTTTACAGAAACACCCCTGGGACACG 1327
QY 441 GlnIleHisArgAspAlaPheSerTyrGlyAlaValAlaGlyProGlnValAspSerArgVal 460
DB 1328 CAGATTCAACGCGATGCTTTTTCGTACGCTGCGCGTCCGTGAGAGTGGACTCTGTGTC 1387
QY 461 IleValAspLeuArgTyrPheGlyAlaThrAspProGluAlaAsnAsnLeuValPhe 480
DB 1388 ATCGTCGACCTGCGCTGGTTGGCGCAACCGACCTGAAAGCAACACTTTTGGTTTTC 1447
QY 481 GlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSer 500

Db 1448 CAGACGATGTTCAAGCGGTACAGTATGCCGACGCCAGCTTCAGATATGACCCAGC 1507
 QY 501 ThrAlaSerAsnValArgAlaArgIlysmetMetAlaAspMetCysGluValAlaSerAsn 520
 Db 1508 ACTGGCTAAACGTAGACGACGAAATGATGGCCGATATGTGGAAATGGCGAGCAAC 1567
 QY 521 LeuGlyGlyThrLeuProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHis 540
 Db 1568 TTGGAGAGTATTTGGCCACGACGCCCGGAGGTATGATGATCCAGGCGCTTGCACTTCAT 1627
 QY 541 LeuAlaGlyThrThrArgIleGlyPheAspIlysalAlaThrThrValAlaAspAsnSer 560
 Db 1628 CTTCGGGGGACTACTCGCATGGCTTCGACAAAGGCAACATACAGTGGCTGATTAACATTCG 1687
 QY 561 LeuValTrpAspPheAlaAsnLeuThrValAlaGlyAsnGlyThrIleArgThrGlyPhe 580
 Db 1688 CTGGCTGGGACTTGGCCATCTTATATGTTCGAGCAATGGCACCATAGACGGGCTTC 1747
 QY 581 GlyIleAsnProThrLeuThrSerMetCysHisAlaIleIlySerAlaArgSerIleIle 600
 Db 1748 GCGGAGAACCCGACACTTACGTCGATGTCGCCAGCGTATCAAGAGCGGAGGAGCATATC 1807
 QY 601 AsnThrLeuIlyGlyGlyThrAspGlyIlyAsnThrGlyIlyHisArgAsnLeu 618
 Db 1808 AATACACTCAAGGGGTGGACTGACGAAATAATACAGCGAGCATCGCAACCTT 1861
 RESULT 2
 AF535193
 LOCUS AF535193 2064 bp mRNA linear PLN 31-DEC-2002
 DEFINITION Peniophora sp. SG pyranose oxidase mRNA, complete cds.
 ACCESSION AF535193
 VERSION AF535193.1 GI:27436421
 KEYWORDS
 SOURCE Peniophora sp. SG
 ORGANISM Peniophora sp. SG
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Aphyllophorales; Lachnocladiaceae; Peniophora.
 REFERENCE 1 (bases 1 to 2064)
 Heckmann, D.M., Bastian, S., Rekowski, M. and Giffhorn, F.
 Pyranose Oxidase of the White-Rot Fungus Peniophora sp. Strain SG: Cloning and Characterization of the Gene, Heterologous Expression in Escherichia coli, and Properties of the Recombinant Enzyme Unpublished
 2 (bases 1 to 2064)
 Heckmann, D.M. and Giffhorn, F.
 Direct Submission
 Submitted (07-AUG-2002) Applied Microbiology, University of the Saarland, Im Stadtwald, 66123 Saarbruecken, Saarland, Germany
 FEATURES
 source
 1..2064
 /organism="Peniophora sp. SG"
 /mol_type="mRNA"
 /strain="SG"
 /db_xref="taxon:204723"
 26..1897
 /EC_number="1.1.3.10"
 /note="FAD-oxidoreductase; Pox; pyranose: oxygen
 2-oxidoreductase; poxSG"
 /codon_start=1
 /product="pyranose oxidase"
 /protein_id="AA013382.1"
 /db_xref="GI:27436422"
 /translation="MSTSSDPFNFNAKSPFSAKASASLPLIPGPKKVGMD
 IKYDVIVGSGPCTCYARELYAGYKRAMDGEIDSGITGAKHKTIVYOKNIDK
 FVNIYOGGLMSVSPVNTLVDTLSPTSMQSTFVRKNSPBDPLNLSGOAVTRV
 VGMSTHMTCAFPFRDREPLLVKDDADADAEWRLYTAESYFOTGTOFESIR
 HNLVNLKAEYKGRDFQOQIPLATRSPFVWSSANVTEDIQNRNTDAPBERFN
 LFPVACERVRNALNSIEISLHIDLSGRFEIKADVYLVLAQNRNTDLVSGF
 GOLGRPNTPNELPLSGSYITFOSLVFCOTVSTELIDVKSOMTIRGREGTYS
 VTTPGASTNKHPMMNKKVNNHMOHEDPLIPDEPPEOVNTLTPRSPHMOIH
 RDAFSYGAVOOSIDSRLLVDMRFGRTRPEKKNIMPSDKITDAVNMOPFDREPPA
 GRISKLEADMTDMCVMSAKIGFLPGSLPQPMEGVLHLHGTHRMKDFEDKENCYV
 NTDSRVFGEFKNLFLGGCGNIPITAYGANTPLAMSLAIKSCYIKONFTPSPTSDAO*

polyA_site 2047
 BASE COUNT 462 a 652 c 579 g 371 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,32e-87
 Score: 129.50
 Percent Similarity: 57.36%
 Best Local Similarity: 43.29%
 Query Match: 37.74%
 DB: 8 Gaps: 16
 US-09-856-327-2 (1-618) x AF535193 (1-2064)
 QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrThrValAlaLeuGluValGlu 56
 Db 167 GACGTTGCATGATGAGCTCCGACCGATGTGATGACAGTATGCCCTGACGCTGCC 226
 QY 57 AlaGlyLeuArgValAlaMetValGluIleGlyAlaAlaAspSerPheThrAlaValAsn 76
 Db 227 GCAGTTACAGAGTGCATGTCGACATCGAGAGATCGACTGTGGCTGAAGATT--- 283
 QY 77 AlaGluGluGlyThrAlaValProThrValProGlyThrHisIlySylsAsnGluIleGlu 96
 Db 284 -----GGTCCCAAGAAAGAACACACCTCGAG 310
 QY 97 PheGlnIlyAspIleAspArgPheValAsnValIleIlySylsGlyAlaLeuGluValSer 116
 Db 311 TACCAAGAACATCGACAGAGTTGTGATGACGATTCAGGGCAATTATATGCTGTTTC 370
 QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaThrSerAlaProPro 136
 Db 371 GTTCCCGCATATACCTCTGCTGTCACACGCTACCCGAGCTGTGGCAAGCT----- 424
 QY 137 GlySerSerAlaIleSerAsnGlyIlyAsnProHisGlnArgGluPheGluAsnLeuSer 156
 Db 425 TCGAGTTCTTCTCCGACACGCGTCGAAACGACGACGCCCTTCGTAACTCAGT 484
 QY 157 AlaGluAlaValThrArgGlyValGlyIlyMetSerThrHisThrThrCysSerThrPro 176
 Db 485 GGTCAAGGGGTACCGCGGTGCTGCGAGCATGTTCGACATGAGCAATGCGCCACACC 544
 QY 177 ArgIleHisProProMetGluSerLeuProGlyIleGlyArgProIlySerAsnAsp 196
 Db 545 CGTTTCGACCGGACGACAA-----CGCGCCCTGCTGTGAGAGAC 583
 QY 197 ProAlaGlu---AspAspIlyGluThrAsnGluLeuThrSerGluAlaGluArgLeuIle 215
 Db 584 GACGCGGACGCTGACGAGCTGAATGGACCGCTACACCAAGCGGAGTCTACTTC 643
 QY 216 GlyThrSerThrIlyGluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeu 235
 Db 644 CAGACTGGACGACCACTTCAGAGATGCAATCCCAACACTCTGCTCAACAGCTC 703
 QY 236 GlnAspAlaThrIlyAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg 255
 Db 704 GCGGAGGAATACAAAGGC---CAACGCACTTCACAGAGATTCACCTGCGCGGACCCGC 760
 QY 256 LeuIlyAsnAlaProGluThrValGluThrPheSerAlaGluAsnLeuPhe-----His 273
 Db 761 -----CGAGCCCGACCTGCTGCAATGAGACTGGGAAACACAGTGTTCAGCTCCAG 814
 QY 274 SerIleThrAsnAspAspIlyGluIlySylsLeuPheThrLeuLeuThrAsnHisArgCys 293
 Db 815 AACAGCCCAACACGAGCGCGCGGAGGAGGCTTCAACCTTCCCGCGGTGCGGT 874
 QY 294 ThrArgLeuAlaLeuThrGlyIlyThrIlySylsIleGlyAlaAlaGluValArgAsn 313
 Db 875 GAGCGCGCT-----GTGGGCAAC 892
 QY 314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerThrYr----- 327
 Db 893 GCGTTG-----AAGTCGAGATCGAGAGTCTGCAATCCAGACCTCATC 937

```

QY 328 -----IleMetAlaLysValTyrValLeuAlaSerGlyAlaIleGly 341
Db 938 TGGGGGACCGCTTCGAAATCAAGGCTGACGTGCTCCTCACCGGGGGGCTGCAC 997
QY 342 AspProGlnIleLeuTyrAsnSerGlyPheSerGlyLeu---GlnValThrProAlaGan 360
Db 998 AACACGCGACTTCTCGAATCTGGCTTGGACACGTGGCGCGCAACCCACCAAC 1057
QY 361 ---AspSerLeuIleProAsnLeuGlyAcyTyrIleThrGluGlnProMetAlaPheCys 379
Db 1058 CCACCGGAGCTGTGGCCGCCCTCGGGGAGCTACATACCGAAGAGTCGCTCTCTGC 1117
QY 380 GlnIleValLeuArgGlnGluPheValAlaSerValAlaAspSerPro-----TyrGly 397
Db 1118 CAGACCGTGATGAGCACCGAGCTACGACAGGTCGAATCCGACATGCATCAGGGGA 1177
QY 398 LeuPro----- 399
Db 1178 ACCCGTGGCGAGCTGACGTACAGCGTACACCGCGCGCGCTGCACCAACAGCAC 1237
QY 400 -----TrpTrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
Db 1238 CCGGACTGGTGGACAGCAAGGTGAATAAACACATGATGACGACGACGAGACCCGCTC 1297
QY 418 ProIleProPheArgAspProGluProGlnValThrThrProPheThrGluGlnHisPro 437
Db 1298 CCGATCCGCTGAGAGACCCCGAGCGGAGGTATACACATCTGTCCACCGCTGCACCG 1357
QY 438 ThrHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAsp 457
Db 1358 TGGCACACTGATGATCCACCGCATGTCTTACGTACGGCGAGTGCAGCAAAACAGCAG 1417
QY 458 SerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnAsnLeu 477
Db 1418 TCGCGCTCATGCTGAGTACGGCGCTTCTTCGGCGGAGCGGAGCCCAAGAGGAAACAAG 1477
QY 478 LeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyr 497
Db 1478 CTCTGGTCTTCAGACAGATTCACGGCGCTACCAACATCCGACGCCACGTTGCGACTTC 1537
QY 498 Arg---ProSerThrAlaSerAsnValAlaArgLysMetMetAlaAspMetCysGlu 516
Db 1538 CCGTTCGCCGGCGCGCGCGCACAGCAAGGCGGAGACATGATGACCGACATGTGCTTC 1597
QY 517 ValAlaSerAsnLeuGlyTyrLeuProThrSerProProGlnPheMetAspProGly 536
Db 1598 ATGTGGCGAAGATGCGTGGCTTCTTACCGCGGCTCCCTCCCGCAATTTCATGGAGCTGCT 1657
QY 537 LeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAla-----ThrThr 554
Db 1658 CTGTGCTCTTACCTTGGTGTACGCGACCCGACATGGGCTTGCAGAGAGAGGACAACTGC 1717
QY 555 ValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGly 574
Db 1718 TGGCTACACAGGACTCGCGCGTTCGCTTCAGAACCTCTCTCTCGGTGGCTGGGA 1777
QY 575 ThrIleArgThrGlyPheGlyLysAsnProThrLeuThrSerMetCysHisAlaIleLys 594
Db 1778 AACATTCTTACCGCGAGCGGCGAAGCCGAGCGCTACCGCAATGTGCTCGCGATCAAG 1837
QY 595 SerAlaArgSerIle 599
Db 1838 AGTTGCGAGTACATC 1852

```

```

SOURCE Trimetes ochracea
ORGANISM Trimetes ochracea
REFERENCE Aphyllophorales: Trimetes.
AUTHORS Vecerek, B., Maresova, H., Kocanova, M. and Kyslik, P.
TITLE Molecular cloning and expression of the pyranose-2 oxidase gene
          from Trimetes ochracea MB49 in Escherichia coli
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2069)
AUTHORS Vecerek, B., Maresova, H. and Kyslik, P.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-2003) Institute of Microbiology AS CR, Videnska
          1083, Prague 4 142 20, Czech Republic
FEATURES
  source
    1..2069
    /organism="Trimetes ochracea"
    /mol_type="mRNA"
    /strain="MB49"
    /db_xref="taxon:230624"
    1..2069
    /gene="p2o"
    49..1920
    /gene="p2o"
    /codon_start=1
    /product="pyranose oxidase"
    /protein_id="AAP40332.1"
    /db_xref="GI:31044224"
    /translation="MSTSSDDPFNFRKSSPRSAANOKASASSLPRLPGDDKKVGM
    IRYDYIVSSGPIGCTIARELYAGCYKVAAMPEIGELDSLKGAKKNTVEOKNID
    FVNVIQGLMSVSVFVNITLVDTLSTSMQASTFVRKNSNEQDPLRLSCQAVTR
    VGGMSTHWTCATPREDEQRLPLVADDAQDAEDRLYTKAESYFQRTDQFKSIR
    HNLVNLKLEEKYGGORDFOQIPLAATRRSPTEVWSSANTVPDLQNRPTDAPBERFN
    LPPAVACERVVKNALNSIEISLHIDILSGDFEELKADYVLTAGAVHTQLVSGF
    GOLGRPNPANPELLPLSGSYLTEOSLYCQVMSSELDVSKDPTIRGTGELTYS
    VYTRPGASNKRHPDKMNEKKVKNHMOHODPLIPREDEPOVTLTFQSPHPMHOIH
    RDAFSGAVQOSIDSRILYDAPRFGRTERKEKNLKFSTIDATNMPPTDFRPA
    GRTSKEDMDMTDCVMSAKTIGFLGSLPQRMPELVALIGTHRMGDFKEDKNCV
    NDSRVFGKNIPLDGGCNIPTAYGAMPILTMASLAIKSCVEYIKONFTPSPTEBAQ"
  BASE COUNT 449 a 659 c 584 g 377 t
  ORIGIN
  Alignment Scores:
    Pred. No.: 2,33e-87 Length: 2069
    Score: 1239.50 Matches: 274
    Percent Similarity: 57.36 Conservative: 73
    Best Local Similarity: 45.29 Mismatches: 173
    Query Match: 37.74% Indels: 85
    DB: 8 Gaps: 16
  US-09-856-327-2 (1-618) x AY291124 (1-2069)
  QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
  Db 190 GACGTTGTATGTAGAGCTCCGACCGGATGGATGACACCTTAAGCCCGTAGCTCGTGGC 249
  QY 57 AlaGlyLeuArgValAlaMetValGluIleGlyAlaAlaAspSerPheTyrAlaValAsn 76
  Db 250 GCAGGTTCACAGGTGCCCATGTTCGACATCGGAGAGATGACTCTGGCCCTGAAGATT--- 306
  QY 77 AlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysLysAsnGluIleGlu 96
  Db 307 -----GGTGGCCACACAGAACACCGTGGAG 333
  QY 97 PheGlnLysAspLleAspArgPheValAsnValIleLysGlyAlaLeuGlnGlnValSer 116
  Db 334 TACCAAGAACATGACAGCAAGTTTGGACGTCATTCAGGGCAACTGATGCTCTTTC 393
  QY 117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaAlaTrpSerAlaProPro 136
  Db 394 GTTCCCGTCATACCTCGTGTGTCGACACGCTCAGCCGAGCTGTGGCAAGCT----- 447
  QY 137 GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGluAsnLeuSer 156

```

```

Db      448 TCGACGTTCTTCGCGCAACGCTCGAACCCAGAGCCACCTTCGTAACTCCAGT 507
Qy      157 AlaGlaValAlaThrArgGlyValAlaGlyMetSerThrHisThrCysSerPro 176
Db      508 GGTGAGGGGTGACGCGGTGCTGAGGACATGTCTACGACTGACATGTGGCCACACC 567
Qy      177 ArgIleHisProPrometGluSerLeuProGlyIleGlyArgProLysLeuSerAsn 196
Db      568 CGTTTCAGACCGCGAGCAA-----CGCCGCTCTCTGTGAGAGAC 606
Qy      197 ProAlaGlu---AspAspLysGluTrpAsnGluLeuTyrSerGluAlaGluArgLeu 215
Db      607 GACCGGACGCTGACGAGCTGAGTGGACCGCTCTACACCAAGCGAGTCTGACTTC 666
Qy      216 GlyHisSerThrLysGluPheAspGluSerIleArgHisThrIleValLeuArgSer 235
Db      667 CAGACTGGACGACGAGCTCAAGATCGATCCGCCAATCTCTGCTCAACAGCTC 726
Qy      236 GluAspAlaTyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHis 255
Db      727 ACGGAGGAATCAAGGCG---CAACGCGACTTCCAGCAGATTCCACTCGCGCAACCC 783
Qy      256 LeuLysAsnAlaProGluTyrValGluTrpHisSerAlaGluAsnLeuPhe-----His 273
Db      784 -----CGAGCGCCGACCTCTGCGAATGAGCTCGCGCAACACTGTTTCGACTCCAG 837
Qy      274 SerIleTyrAsnAspAspLysGluLysLysLeuPheThrLeuLeuThrAsnHisArg 293
Db      838 AACAGGCGGAACGACGCGCGCGAGGCGCTTCAACCTTCCCGCGGTGCGCGT 897
Qy      294 ThrArgLeuAlaLeuThrGlyGlyTyrGlyLysLysIleGlyAlaAlaGluValArg 313
Db      898 GAGGCGCGT-----GTGCGCAAC 915
Qy      314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyr----- 327
Db      916 GCGTTG-----AACTCGGAGATCGAGATCTGCACATCCAGACCTCATC 960
Qy      328 -----IleMetAlaLysValTyrValIleuAlaSerGlyAlaIleGly 341
Db      961 TCGGGCGACCGCTTCGAATCAAGCTGACCTGTACCTGCACCGCGGGCGCTGCAC 1020
Qy      342 AsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeu---GlnValThrProArg 360
Db      1021 AACACGCGACTTCTCGTAACCTGTGCTTGGACAGCTGGCCCGCAACCCGCCAAC 1080
Qy      361 ---AspSerLeuIleProAsnLeuGlyArgTyrIleThrGlnGlnProMetAlaPhe 379
Db      1081 CCACCGGAGCTGCTGCGTCCCTGGGAGTACATCACCGACGAGTCTGCTCTGTC 1140
Qy      380 GlnIleValIleuAlaGlnGluPheValAspSerValAlaArgAspPro-----Tyr 397
Db      1141 CAGACCGTGATGACACCGACTCATCGACGCTCAAGTCCGACATGACCATCAGGGA 1200
Qy      398 LeuPro----- 399
Db      1201 ACCCTGGCGGACGTACAGCTGACGTACACGCGGGCGCTGCACCAACAGCAC 1260
Qy      400 -----TrpTyrLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAsp 417
Db      1261 CCGGACTGTGTAACGAGAAAGTGAATAAACACATGATGACAGCACGAGAGACCG 1320
Qy      418 ProIleProPheArgAspProGluProGlnValThrThrProPheThrGluGluHis 437
Db      1321 CCGATCCCGTGGAGCCCGAGCGCGAGTTCACCTCTGTTCAACCGTCAACCGG 1380
Qy      438 TrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluVal 457
Db      1381 TGGGACACTGATCCACCGCGATGCTTTCAGTTACGGCGAGTCCAGCAAGCATC 1440
Qy      458 SerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnLeu 477

```

```

Db      1441 TCGGCTCATCTGAGCTGCGCTTCTTCGCGCGAGAGCCCAAGAGAGAACAG 1500
Qy      478 LeuValPheGluAsnAspValAlaGlnAspGlyTyrSerMetProGlnProThrPhe 497
Db      1501 CTCTGCTTCTAGACAATACAGACCGCGTACAAACATCCCGACCGCATCTTGACTTC 1560
Qy      498 Arg---ProSerThrAlaSerAsnValArgAlaArgLysMetMetAlaAspMetCys 516
Db      1561 CGCTTCCCGCGCGCGCGAGCAGACGAGCGGAGGACATGATGACCGACATGCGCTT 1620
Qy      517 ValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProProGlnPheMetAspPro 536
Db      1621 ATGTGCGGGAAGATCGGGCTTCTTACCCGCTCCCTCCGCAATTCATGAGCTGTG 1680
Qy      537 LeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAla-----Thr 554
Db      1681 CTGTGCTTCACTTGTGTGTGATGACCGACCGCATGGCTTCGACGAGAGAGACACTGC 1740
Qy      555 ValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnLeuTyrValAlaGly 574
Db      1741 TGGCGCAACAGAGACTCGCGGTGTGCGCTTCAAGAACCTTCTCGTGGTGGCGGA 1800
Qy      575 ThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAla 594
Db      1801 AACATCTCTACCGGTACGCGCGCAACCGACGCTCACCGGAATGTCTGCGATCA 1860
Qy      595 SerAlaArgSerIle 599
Db      1861 AGTTGCGAGTACATC 1875

RESULT 4
LOCUS      E11766                      1869 bp    RNA       Linear    PAT 29-SEP-1997
DEFINITION E11766, cDNA of pyranose oxidase.
ACCESSION  E11766.1 GI:22025402
VERSION    JP 1996205861-A/1.
KEYWORDS   pyranose oxidase
SOURCE     unclassified
ORGANISM   unclassified.
            (bases 1 to 1869)
REFERENCE  1 (bases 1 to 1869)
            Nishimura,I., Okada,K., Nanbara,T., Kawai,G., Koyama,T. and
            Suzuki,M.
            NEW PYRANOSE OXIDASE, PYRANOSE OXIDASE GENE, NEW RECOMBINANT DNA
            AND PRODUCTION OF PYRANOSE OXIDASE
            Patent: JP 1996205861 A 1 13-AUG-1996;
            KIRKMAN CORP
COMMENT    OS      Coriolus versicolor
            PN      JP 1996205861-A/1
            PD      13-AUG-1996
            PE      24-MAY-1995 JP 1995124835
            PR      07-DEC-1994 JP 94P 304086
            PI      NISHIMURA IKUKO, OKADA KIMI HARU, NANBARA TOMOYUKI, PI
            KAWAI
            GENSUIRO,
            PI      KOYAMA TAIJI, SUZUKI MASARU
            PC      C12N9/04,C12N15/09,(C12N9/04,C12R1.19),(C12N15/09,C12R1.01);
            CC      strandedness: Double;
            FH      topology: Linear;
            FT      key
FEATURES   source      1..1869
            /strain='ps4a',
            location/Qualifiers
            1..1869
            /organism='unidentified'
            /mol_type='genomic RNA'
            /db_xref='taxon:32644'
BASE COUNT  408 a      593 c      529 g      339 t
ORIGIN
Alignment Scores:      7.12e-87      Length:      1869
Pred. No.:

```


process for producing pyranose oxidase
Patent: US 5712139-A 1 27-JAN-1998;

JOURNAL
FEATURES
source 1. 1869
Location/Qualifiers

BASE COUNT 408 a 593 c 529 g 339 t
ORIGIN

Alignment Scores:

Pred. No.:	7 126-87	Length:	1869
Score:	1232.50	Matches:	273
Percent Similarity:	57.00%	Conservative:	73
Best Local Similarity:	44.98%	Mismatches:	176
Query Match:	37.53%	Indels:	85
DB:	6	Gaps:	16

US-09-856-327-2 (1-618) x 182410 (1-1869)

```

QY 37 AspValpHeilelaaglyserglyproileglyalathrTyralaLysleucysValgIu 56
    |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 142 GACGTGTTCATAGTAGGCTCGGACCGAATGAGTACAGTATGCCCGTACGCTCGCA 201
QY 57 AlaGlyLeuArGValaValmeValaIuileglyAlaAlaAspSerPheTyraValaLsn 76
    |||||  |||||  |||  |||  |||||  |||||  |||||  |||||  |||||  |||||
Db 202 GCCGGTTACAAGGTCGCCATGTTCCACATCGGGAAATTAATGACTCGGCTGAAGATC--- 258
QY 77 AlaGluGluGlyThraValaProtyrValaProglyTyraHisLysLysasnGluLeu 96
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 259 -----GGTGGCCCAAGAGAACACCGCTCGA 285
QY 97 PheGlnLysAspIleAspArgPheValaLsnValaIleLysGlyAlaLeuGlnGlnValSer 116
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
Db 286 TACAGAGAACATGATGACAGTTGTGAACGTCATTCAGGCCCAATGATGTCGTTC 345
QY 117 ValProValaArgasnGlnAsnValaProthraLeuAspProglyAlaTrpSerAlaProPro 136
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 346 GTTCCGTCATACCTCGTGTGATCGACACCGACCGACGCTGTGGCAACCT----- 399
QY 137 GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGlnAsnLeuSer 156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 400 TCATGCTTCCTGTCGCGAATGGCTCGAACCAGACGAGACCGCTTCGTAACCTCACT 459
QY 157 AlaGlnAlaValaThraArgGlyAlaGlyLysMetSerThraHisTrpThrCysSerThraPro 176
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
Db 460 GGTCAAGGGGTCACCGCTGTCGTGAGGACATGTCACGACATGACATGCCGACACCG 519
QY 177 ArgIleHisProIleMetGlnSerLeuProGlyIleGlyArgProLysLeuSerAsnsp 196
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 520 CGCTTTGACCGGACGACG-----CGCCCTGCTGCTGTAAGGAC 558
QY 197 ProAlaGlu-----AspAspLysGluTrpAsnGluLeuTyraSerGlnAlaGlnArgLeuIle 215
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 559 GACCAAGACGCTGACGACGCGAGTGGAGCGGCTGTACACCAAGGCGGAGTCATCTTC 618
QY 216 GlyThrSerThrLysGluPheAspLysIleArgHisIsthLeuValaLeuArgSerLeu 235
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 619 AAGACCGGAGCGACGATCAAGAGATCGATCCGCAACACCTGCTGTCAACAACGTC 678
QY 236 GlnAspAlaIlyrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg 255
    ::|||  ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 679 GCGGAGGAATCAAAAGT---CAGCGCAACCTTCAGAGATCCCTCGCGGCAACGCGT 735
QY 256 LeuLysAsnAlaProGluTyraValaGluTrpHisSerAlaGlnAsnLeuPhe-----His 273
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 736 -----CGCAGTCCGACCTTCGTCAGTGGAGCTCGGCAACACCGTGTTCACCTCCAG 789
QY 274 SerIleTyraAsnAspLysGlnLysLysLeuPheThrLeuLeuThraAsnHisArgCys 293
    ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 790 AACAGCGGACGACGCGCGGAGATGAGCGCTTCAACCTCTCCCGCGGTGCATGT 849
QY 294 ThrArgLeuAlaLeuThraGlyGlyTyraGlyLysLysIleGlyAlaAlaGluValaArgasn 313
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 850 GAGCGCGTC-----GTGGCAAC 867

```

```

QY 314 LeuLeuAlaThraArgAsnProSerSerGlnLeuAspSerTyra----- 327
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
Db 868 -----ACGTGCAACTCCGAGATCGAGAGCTTCACATCCACGACCTCATC 912
QY 328 -----IleMetAlaLysValTyraValaLeuAlaSerGlyAlaIleGly 341
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 913 TCGGGCAGACCGCTTCGAATATCAAAAGCAGACGTCGTCTTCAACAGCGGGCGGTCCAC 972
QY 342 AsnProGlnIleLeuTyraAsnSerGlyPheSerGlyLeu---GlnValThraProArgasn 360
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 973 AACGGCAGACCTTCGTGAACCTGCTTGGACAGCTGGCGCGGACCGCGAGAAC 1032
QY 361 ---AspSerLeuIleProAsnLeuGlyArgTyraIleThrGlnGluProMetAlaPheCys 379
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1033 CCGCGCGACCTTCGCTCGCTCCCTCGGAACTACATCACCGACACCTGCTCTTCTTC 1092
QY 380 GlnIleValaLeuArgGlnGluPheValaLsnSerValaArgAspAspProTyra-----Gly 397
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1093 CAGACCGGTATAGCAGCAGAGCTCATCGACGTCATCAAGTCCGACATGATCATCAGGGGC 1152
QY 398 LeuPro----- 399
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1153 AACCTGGCGATCTGGGTACAGCGTCACGTCACGCGCGCGGAGACCAACAGACAC 1212
QY 400 -----TrrPrLysGluAlaValaIleGlnHisIleAlaLysAsnProThraSpAlaLeu 417
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1213 CCGGACGTGTGAGAACGAAAGGTGAAGAACCAATGATGACACACGAGAGACCGCTT 1272
QY 418 ProIleProPheArgAspProGluProGlnValaThraThrProPheThrGlnLysHisPro 437
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1273 CCAATCCGCTTGAGAGACCCGAGCGGACGACACCTTCTTCAGCCATCGACGACCG 1332
QY 438 TrpHisThrGlnIleHisArgAspAlaPheSerTyraGlyAlaValaGlyProGluValaAsp 457
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1333 TGGCAGACTCAATTCACCGGATCGCTTCATTCAGGCGGCTGACGCAAGCATCGAC 1392
QY 458 SerArgValaIleValaLysLeuArgTrpPheGlyAlaThraAspProGluAlaAsnLeu 477
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1393 TCACGCTTCATGTCGATCGCGGCTTCCTTCGCGCGGACGAGACCAAGAGAGAACAAAG 1452
QY 478 LeuValPheGlnAsnAspValaGlnAspGlyTyraSerMetProGlnProThraPheArgTyra 497
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1453 CTGCTGCTTCGCAAAATATACGACACGATACATGCGCACCGCGCTTCGACTTC 1512
QY 498 Arg---ProSerThraLysAsnValaArgAlaArgLysMetMetAlaAspMetCysGlu 516
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1513 CGCTTCGCGGCGCGGACGACGAGGAGGAGGACATGATGACCGATATGTGCGCTT 1572
QY 517 ValAlaSerAsnLeuGlyLysLeuProThraSerProProGlnIlePheMetAspProGly 536
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1573 ATGTGGCGAGATGTGGCTTCCTGCTCCGCGCTCCCTCCCAATTCATGAGCCCGGT 1632
QY 537 LeuAlaLeuHisLeuAlaGlyThraArgIleGlyPheAspLysAla-----ThrThr 554
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1633 CTGTGCTTCACCTCGGTGTGATCGCAGCAGATGCTTCGAGCAGAGAGACAAAGTGC 1692
QY 555 ValAlaAspAsnAsnSerLeuValaTrpAspPheAlaAsnLeuTyraValaIleLysnGly 574
    ::|||  ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1693 TGGCTGACACAGCTCGCGCTGTGGCTTCAAGAACCTTCCTCGTGGTGGCGGA 1752
QY 575 ThrIleArgThrGlyPheGlyLysnProThraLeuThraSerMetCysHisAlaLys 594
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1753 AACATTCACCGGCTACGCGCGGACCAACCGACGCTACCGCAATGTCTGCGCATCAAG 1812
QY 595 SerAlaArgSerIleLeasn 601
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1813 AGTTCGAGTACATCAAGAAC 1833

```

RESULT 6
LOCUS D73369 2046 bp mRNA linear PLN 04-FEB-1999
DEFINITION Corioliu versicolor mRNA for pyranose oxidase, complete cds.

QY	398	LeuPro-----	399
Db	1178	AAACCTGGCGGACATCTGGGGTTACAGCGTCACGTACAGCCGGCGGAGGAAACCAACAGCAC	1237
QY	400	-----TTPTPlySGluAlaValAlaIaInHisIleAlaIyAsnProThrAspAlaIeu	417
Db	1238	CCGACGTGGTGAACGAAAGGTGAAGAACCAATGATGACAGACACAGAGACCCGGTT	1297
QY	418	ProIleProPheArgAspProGluProGlnValThrThrProPheThGluGluHisPro	437
Db	1298	CCAAATCCGGTTGAGAGACCCGAGCGGAGGTACACACTGTTGTCACACCATCGACCCG	1357
QY	438	TriPHisThrGlnIleHisArgAspAlaPheSerTrgIyAlaValGlyProGluValAsp	457
Db	1358	TGGCACACTCCAGATTCCACCGGATCCGTTACGTCAGCGCGGTGCACAAAGCTGCAC	1417
QY	458	SerArgValIleValAspLeuArgTrpPheGlyAlaThrAspProGluAlaAsnIleu	477
Db	1418	TCACGTCTCATCTGCTGACTGGCGCTTCTTCGGCGGAGGAGCCAAAGAGAGACAAG	1477
QY	478	LeuValPheGlnAsnAspValGlnAspGlyTrpSerMetProGlnProThrPheArgTrp	497
Db	1478	CTCTGCTCTCGGACAAATTAACGACACGTAACAACATGCCGACCGAGCTGCACCTC	1537
QY	498	Arg---ProSerThrAlaSerAsnValArgAlaArgIySmetMetAlaAspMetCysGlu	516
Db	1538	CGCTTCCCGCGGCGCCGACGAGGAGGCGGAGCATATGACCATATGTGCTT	1597
QY	517	ValAlaSerAsnIleuGlyIyTrpIleuProThrSerProProGlnPheMetAspProGly	536
Db	1598	ATGTGGCGGAAGATTGTGGCTCTCTCGCCCGGCTCCCTCCGCAATTTATGAGCCCGGT	1657
QY	537	LeuAlaIeuHisIleuAlaGlyThrThrArgIleGlyPheAspIyAla-----ThrThr	554
Db	1658	CTTGCTCTTCACCTCGGTGTGTAACGACGCGCTTCGACGACGACGAGCAAGCAATGTC	1717
QY	555	ValAlaAspAsnAsnSerLeuValTrpAspPheAlaAsnIleuTrpValAlaGlyAsnGly	574
Db	1718	TGCGTCAACCGGACTCCCGCGCTTGTGGCTTCAAGAACCTGTCTCGTGGTGGCGGA	1777
QY	575	ThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAlaIleIyS	594
Db	1778	AACATTCACCGCGGTGAGGCGGACCCGACGCGTACACCGCATGTGCTGCGCATCAG	1837
QY	595	SerAlaArgSerIleIleAsn	601
Db	1838	AGTTGCGAGTACATCAAGAAC	1858
RESULT 7			
E32737	LOCUS	E32737	1902 bp DNA linear PAT 18-JUN-2001
DEFINITION	Panus cornucopiae-derived antitumor protein and gene thereof.		
ACCESSION	E32737		
VERSION	E32737.1	GI:13017467	
KEYWORDS	JP 1999315096-A/1.		
SOURCE	Pleurotus cornucopiae		
ORGANISM	Pleurotus cornucopiae (cornucopia mushroom)		
REFERENCE	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;		
AUTHORS	Agaricales; Pleurotaceae; Pleurotus.		
TITLE	Akhihiro,M., Koji,I., Kazuaki,K., Tomohide,S. and Ryoko,A.		
JOURNAL	Panus cornucopiae-derived antitumor protein and gene thereof		
COMMENT	Patent: JP 1999315096-A 1 16-NOV-1999;		
	NYUFUDO KURIESHON GIJUTSUOKENYUKUMIAI		
	OS Pleurotus cornucopiae		
	PN JP 1999315096-A/1		
	PD 16-NOV-1999		
	PF 07-AUG-1998 JP 1998236349		
	PR		
	PI AKIHIRO MORIYA,KOJI IZUMO,KAZUAKI KIDO,TOMOHIDE SAKA, PI		
	RYOKO AOKI		
	PC C07K14/375,.A61K35/84,.A61K38/00,.C07K16/14		

	CC	Key	Location/Qualifiers
FEATURES	FT	source	1..1902
source	FT	location/Qualifiers	/organism="Pleurotus cornucopiae".
BASE COUNT	437 a	556 c	479 g 430 t
ORIGIN			
Alignment Scores:			
Pred. No.:	7.99e-87	Length:	1902
Score:	1232.00	Matches:	275
Percent Similarity:	57.05%	Conservative:	77
Best Local Similarity:	44.57%	Mismatches:	193
Query Match:	37.52%	Indels:	72
DB:	6	Gaps:	18
US-09-856-327-2 (1-618) x E32737 (1-1902)			
QY	18	11leasnglyglnlilerolysasna1a1lehs1gltlthrttyrcllysasnspglyvalasp	37
DB	130	ATTCGGGGAAACCTTCGCGGGA-----GATCAATCAAGATTCGAC	171
QY	38	ValPhe1leAlG1ySerG1yPro1leg1yAlAThTyrAlaLysLysLysValG1uAla	57
DB	172	GTTGTCATCGTGGCGCTGCGCCAGTTGGCTCTACTTATCCCGTCTCCGTCGAGCA	231
QY	58	glyleuavVala1weValG1u1leg1yAla1a1a1aspSerPheTyrAlaVala1asn1a	77
DB	232	GTTTCCAAGTGAAGTATGTTGAAATTTGAGAAATCGACTCT-----	273
QY	78	gluglulg1yTh1a1a1a1ProTyrVal1ProG1yTTh1s1yLysasn1u1leg1uPhe	97
DB	274	-----GGCCGGAAGCTT-----GGTCCCATTAAGAAATACGGTTGACTAC	315
QY	98	g1n1yasp11e1asp1rPheVala1asnVal11e1yG1yAla1eug1ng1a1seVal	117
DB	316	CAGAAAGACATTGACAAATTCGTCATTCATTCAGAGCACTCATGCTGTATCCGTC	375
QY	118	ProVal1ArgasnG1na1naVal1ProTh1e1u1a1sp1roG1yAla1a1p1seVal1a1a1	137
DB	376	CCTGTCAACAATATGTTCCGATACTCTACAGCCCGGTAGCGGCAAGGCC-----TCA	429
QY	138	SerSerAla11e1SerasnG1yLysasnPro1h1s1n1a1a1g1u1Phe1u1a1n1e1u1seVala	157
DB	430	ACACACTTTGTCGCGAAGCCGCAAAACCCGAGCAAGATCCCTTACTATCTCTCAGGC	489
QY	158	g1uAlaVala1Th1cArg1yVala1g1yG1yMetSerTh1h1s1rPh1cY1s1e1rTh1ProArg	177
DB	490	CAGGCTGATCCCGCTTGTGGAGGAATGGCAGCAGCATGGACTTGGCAGCAACCTGCG	549
QY	178	11e1h1s1Pro1MetG1u1se1r1eu1ProG1y11e1g1y1Arg1Pro1y1s1eu1--Ser1asn1asp	196
DB	550	TTCACACAGAGT-----GACGGCGCTAAGCTTGTGAAGAATGAT	588
QY	197	ProAlaG1u1a1sp1a1s1y1s1u1Tr1pa1snG1u1e1u1Tyr1SerG1u1a1a1u1a1g1e1u11e1g1y	216
DB	589	GACTCCGGGAGCCATCGATCGATGAAGCAACCTCTATGACATCCGCCAATCGTTCGTCAG	648
QY	217	Th1SerTh1y1s1G1u1Phe1a1s1r1e1u1a1n1e1h1s1Th1e1u1a11e1u1a1y1se1r1e1u1n	236
DB	649	ACGGGGCATATACGTTGATCAATATATTCGCGCAACCTTACTCTCGAAGCTTCAG	708
QY	237	Asp1a1a1y1y1Lys1a1p1a1rG1n1a1rG11e1Phe1a1rP1ro1e1u1a1a1c1y1h1s1Arg1e1u	256
DB	709	CAGTCTCTATGACAGT---CAGCGCGGATTTGAGACAAATCCCTTTCGCGGCCCAACGAGCC	765
QY	257	Lys1asn1a1a1ProG1u1y1Val1u1Tr1Phe1s1e1r1a1a1g1u1asn1e1u1Phe-----His1ser	274
DB	766	AAC-----CCCCGCTGTGAGTGAAGTCTTGGCCATACGGCTTTTGACTTGAAGAC	819

OY		275	IlelyrAsnAspRpyLysGLNLYSLysLeuHeThrlLeuThrAsnHisAlGcSynrh	294
Db		820	CGCCCAAAAGCGGGATGACAGAAAGGGCGTTTCAACTCTTCGCCGACGTGTATGTGA	879
OY		295	ArgLeuAlaLeuthrNgLyLTyrGlULSLysIleGlYalaLaGluValArgAsnLeu	314
Db		880	CGCCTTAGC---CGGGATTCCCTGCAGCGAATAGAGCATTAAGGTGATCATGACCTG	936
OY		315	LeuAlaThrArAsanProSerSercLIneuAspSerTr---lleMeAlaLYValTYrr	333
Db		937	ATTAGCGC-----GATGCTRAAGAAGTCAAGCGCTGACGTGTT	975
OY		334	ValLeuAlaSerGlyYalAleGlyAsnProGlnlleLeuTyraSnSergLyPheSercLy	353
Db		976	ATCCTTGTCCTGCGCGAGCTCCATAATCCCCAGATTCTGTAAACTGTGATTC---GGG	1032
OY		354	LeuGlnValThrProAlgAsnAsp-----SerLeulleProAnLeuGlyYarg	369
Db		1033	CGCATGGGTGACCGAGATTCTTCACTGCGCACGCCGCACTCTGCTGCTATCTGTAGGC	1092
OY		370	TyrllethrNgLuGlnPrometalAPhecysGlnlleValLeuArGngLnUPheValAsp	389
Db		1093	TACATCACTAGACAGACACTACGTTCTGCCAGAACCGTTTCAGCACCGAACTGTCAAT	1152
OY		390	SerValArGpaSpasPro----TyrGlyLeuPro-----	399
Db		1153	CTTGTCAAAGTCGGACATGATATTATTGTGGCACGCCGCCGACAGCCAGACTACAGTGTACT	1212
OY		400	-----TrrPrLySGluAlaValAlaGln	407
Db		1213	TTCAACCCCGGACAGCTCCGAGCAATAGACCACCAAAATGTGTGGAAAGAGAGCTACAGAG	1272
OY		408	HslleAlalySaSnProThrAsPaLaLeuProIleProPheArGaSpProGluProGln	427
Db		1273	CACATGATGACGACACAGAGAGATCCCCTCCGATCCGGCTCGATAGACCTCGAGCCACAG	1332
OY		428	ValIthrThrProPhethrNgLuHIsProTrphIsThrGlnlleHlsArGaSPAlaPhe	447
Db		1333	GTCAGCAAGCTCTTCCAGAGCACACATCCATGAGCAGACTCGATTCAATCGTAAGCCCTT	1392
OY		448	SerTyrgLyAlaValGlyProGluValAlaSpSerArqVallleAlaAspLeuArTrPhe	467
Db		1393	AGTCAAGCGCGCTGTGGGGAGAGATAGTGCAGCCGCTGTGTtTGAOCTGGCGCTTCTTC	1452
OY		468	GlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGlnAspGly	487
Db		1453	GGTGGCACTGAACTGTGGAGAGACAAGCTGTGGTTCCTCAAAAAGATCACTGATGCA	1512
OY		488	TyrsErmetProGlnProThrPheArGtyrArg---ProserThrlAlaserAsnValArg	506
Db		1513	TTCAAACCTGCCACAGCCCACTTAACTTCGCGTTCCCCACAGGGCCGACAGCCACAGAG	1572
OY		507	AlaArGlyMetAlaAspMetCysGluValAlaSerAsnLeuGlyLTyrLeuPro	526
Db		1573	GCCGAGCTCATGATGGCTGATGTCACAGTGCACCAAAAGTCGGGGCTTCCTGCC	1632
OY		527	ThrsErProProGlnPheMetAspProGlyLeuAlaleuHIsleuAlaGlyTYrThrArg	546
Db		1633	GGTTCAATATCCGCAAGTTTTATGGCTCTCGGCTTGTACTCACCTTGtGGAAACCCACCG	1692
OY		547	IleGlyPheAspLySalatThrThValAla-----AspaSnAsnSerLeuValTrPasp	564
Db		1693	ATGGGGCTTGATGAGGAAGCTGATAGAAGCGTGTCTGCACACCAACTCAAAAGTCTTGGCC	1752
OY		565	PheaAlaSnLeuTyryValAlaGlyAsnGlyThrIleArGtyrNgLyPheGlyGluAsnPro	584
Db		1753	ATGGAAGAACCCTGTTCTGGCGGTTGGGGCAAACTTGGCAACCGGCTATGCTCGAACCOCG	1812
OY		585	ThrlenuhrSerMetCysHISAlaIlleLysSerAlaArgSerllelleasn	601
Db		1813	ACGCTCACTGGGGTGGCGTTCGACATAGAGAGCTGTAAGTATATACAGAAAT	1863

LOCUS	ARL41573	1869 bp	DNA	linear	PAT 08-AUG-2001
RESULT 8					
ARL41573					
DEFINITION	Sequence 1 from patent US 6146865.				
ACCESSION	ARL41573				
VERSION	ARL41573.1				
KEYWORDS	GI:15101089				
ORGANISM	Unknown.				
REFERENCE	Unknown.				
AUTHORS	Unclassified.				
TITLE	1 (bases 1 to 1869)				
JOURNAL	Christensen, S., Lassen, S., Flemsted, and Schneider, P.				
FEATURES	Nucleic acids encoding polypeptides having pyranose oxidase activity				
SOURCE	Patent: US 6146865-A 1 14-NOV-2000;				
BASE COUNT	Location/Qualifiers				
ORIGIN	1..1869				
	/organism="unknown"				
	397 a 622 c 548 g 302 t				
Alignment Scores:					
Pred. No.:	4.09e-85	Length:	1869		
Score:	1210.00	Matches:	274		
Percent Similarity:	56.19%	Conservative:	71		
Best Local Similarity:	44.63%	Mismatches:	197		
Query Match:	36.85%	Indels:	72		
DB:	6	Gaps:	16		
US-09-856-327-2 (1-618) x ARL41573 (1-1869)					
QY	37 AspyAlPheIlleAlaGlySerGlyProIleGlyAlaThrTyraLaLysLeuCyValGlu	56			
DB	142 GACGTTGCATCGTCGGCTGCGGCCGATGTGCGCATATWCGCCGCGACTGTGAG	201			
QY	57 AlAGlyLeuArgValMetValGluIleGlyAlaAlaSerPheTyraLaValAsn	76			
DB	202 GCCGCGTTCAACGTGCCATGTTCGATGGAAGATCGACTCCGGCTGAGATC--	258			
QY	77 AlaGluGluGlyThrAlaValProTyraValProGlyThrHisLysLysAsnGluIleGlu	96			
DB	259 -----GGCTCACAGAAAGAACACCGCTGAG	285			
QY	97 PheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnValSer	116			
DB	286 TACACAGAAACATCGACAAATTCGTAATGTATACAAAGGCAACTATGCCCGTCTCG	345			
QY	117 ValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaTyrPserAlaProPro	136			
DB	346 GTGCCCGTCACACAGATGTGTGTGACACGTAGACCGCGGCGTATGGCAAGCT-----	399			
QY	137 GlySerSerAlaIleSerAsnGlyLysAsnProHisGlnArgGluPheGlnAsnLeuSer	156			
DB	400 TCGACGTTCTTCGTCGCCGACGGGGCGCATTCACAGCAAGACCCGCTGCCGACCTTAGT	459			
QY	157 AlaGluAlaValThrArgGlyValGluGlyMetSerThrHisThrPThrCysSerThrPro	176			
DB	460 GCCACAGCGGTGCACCCGCGTGTGGCGGCGATGTCTACGCGACTGAGCGCGACCCCG	519			
QY	177 ArgIleHisProPromeGluSerLeuProGlyIleGlyArgProLysLeu--SerAsn	195			
DB	520 CGCTTCGAGAG-----CTGACAGCGCGCGTCTGTCGTAAGAAC	558			
QY	196 AspProAlaGluAspAspLysGluTTPraSngluLeuTyrsSerGluAlaGluArgLeuIle	215			
DB	559 GACATCCAGGCGGCGACGCGCGAGTGGGACAGGCTTACACAAAGAGCGCGAGTCTATTC	618			
QY	216 GlyThrSerThrLysGluPheAspGluSerIleArgHisThrLeuValIleArgSerLeu	235			
DB	619 AAGACCGGCGACGCCACTTGTGCGGAGTCCGACATCCGACCAACCTCGTGTACAAACCTG	678			
QY	236 GlnAspAlaTyrrLysAspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArg	255			

Db 679 CAGGAGAGTACAAGGC---GTGCGGCACTTCCAGAGATCCGCTCGCGGCGAGCGGC 735
 QY 256 LeuysanlaProgluYrValIGluThrHisSerAlaGluAsnLeuPhe-----His 273
 Db 736 -----CAGAGCCCGAGCTTGTGAGTGGAGCTGGGCGCACACCGCTGTGATCTCGAG 789
 QY 274 SerLeuYrAsnAspArgLysGlnLysLysLeuPheThrLeuThrAsnHisArgCys 293
 Db 790 AACCGGCGGAAACAAGAGCGGCGAGAGAGCGCTTCAACSTCTTCCCGCGGCTGGCGTGC 849
 QY 294 ThrArgLeuAlaLeuThrGlyGlyTyrGlyLysLysIleGlyAlaAlaGluValArgAsn 313
 Db 850 ACGAGCTGAGCGCGCATACGCGACSTCGAGATCGTAGCG---CTGGAGTCCGCGGAC 906
 QY 314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAlaLysValTyr 333
 Db 907 CTCACGCGGGGCAAGACATCAC---ATCAAGCGCAAGGTTGAC 948
 QY 334 ValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGly 353
 Db 949 ATCTCAACCGCGCGGCGGTCCACACAGCGCGAGCTCTCGCGGCTGTGGATTGGGCGAG 1008
 QY 354 LeuGlnValThrProArgAsnPro-----SerLeuIleProAsnLeu 367
 Db 1009 CTG-----GGTCGTCCCGACCGCGCAAGCGGCTGCGCTCTGCTGCTCGTACTCTG 1059
 QY 368 GLYArgTyrIleThrGluGlnProMetAlaPheCysGlnIleValLeuArgGlnGluPhe 387
 Db 1060 GGGACCCACATCGACGAGACAGCGCTGTCTTGCCAGCGTCATGAGCAGCAGAGCTC 1119
 QY 388 ValAspSerValArgAspPro-----TyrGlyLeuPro----- 399
 Db 1120 ATCAACAGTGTACACCGGAGATATGACCATCTGTCCGCAAGCCCGGACCCGAGCTATAGC 1179
 QY 400 -----TrrTrrLysGluAlaVal 405
 Db 1180 GTCACTATATACCCCGGCAACCCGACACACAGCACCCGAGCTGTGGAGACGAGAAGTG 1239
 QY 406 AlGlnHisIleAlaLysAsnProThrAspAlaLeuProIleProPheArgAspProGlu 425
 Db 1240 AAGAAGCATGATGAGCACCCAGAGAGACCCGCTCCGATCCGCTCGAGAGCCCTGAG 1299
 QY 426 ProGlnValThrThrProPheThrGluGlnHisProTrrHisThrGlnIleHisArgAsp 445
 Db 1300 CCGAGGTCAACAGCGCTTTCAGAGCAGCAGCACCAGTGGCGCACCCAGATTCCACCGGAC 1359
 QY 446 AlaPheSerTyrGlyAlaValGlyProGluValAspSerArgValIleValAspLeuArg 465
 Db 1360 GCCTTGAGTACGGCGCGTGCACAGACATCGATCGCGCTCATGTGCTGACTGGCGG 1419
 QY 466 ThrPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspValGln 485
 Db 1420 TTCTTCGAGCGACCGAGCCCAAGAGAGACAAAGCTATGCTTCTGGAGACATGACG 1479
 QY 486 AspLysTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsnVal 505
 Db 1480 GAGCGGTACAACCTCCGAGACCGACGTCGACTTCGCG-----TTCCCGCGGGGGCGCC 1533
 QY 506 ArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGlyTyrLeu 525
 Db 1534 GAAGCGGAGAGCATGATGACCGCATGTGCTCATGTGCGGAGAGATCGGTGGATTCTCG 1593
 QY 526 ProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHisIleAlaGlyThrThr 545
 Db 1594 CTTGGGTCTTACCCACAGTTTCATGAGCCGCTTCTGCTGCACTTGGTGGAGACGAC 1653
 QY 546 ArgIleGlyPheAspLysAlaThr-----ThrValAlaAspAsnAsnSerLeuValTyr 563
 Db 1654 CGCATGGCTTCGAGAGAGCGGACAGTGTGCTGCTGCAACGACTCAGCGGTCTTC 1713
 QY 564 AspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGluAsn 583
 Db 1714 GCGTTCAAGAACCTTCTCTGCGGCTGGGGAACATCCACCGGTACGCGCGCAAC 1773

QY 584 ProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleIleAsnThrLeu 603
 Db 1774 CCGAGCGTCAACCGGAAATGCTGCTTGCATCAAGAGCTCGAGTACATC-----AAG 1824
 QY 604 LysGlyGlyThrAspGlyLysAsnThrGlyGlnHisArgAsn 617
 Db 1825 AAGAAGCTCGAGCGGAGCGGACCCGATCGGTGAAGACACCAAC 1866
 RESULT 9
 ARI70485
 LOCUS ARI70485 1701 bp DNA linear PAT 17-DEC-2001
 DEFINITION Sequence 2 from patent US 6291648.
 ACCESSION ARI70485
 VERSION ARI70485.1 GI:17908444
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1701)
 AUTHORS Kawamura,Y., Morita,A., Izumo,K. and Saka,T.
 TITLE Antitumor protein and corresponding gene sequence isolated from
 matsutake mushrooms
 JOURNAL Patent: US 6291648-A 2 18-SEP-2001;
 FEATURES location/Qualifiers
 source 1..1701
 BASE COUNT 411 a 463 c 432 g 395 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7,24e-76 Length: 1701
 Score: 1091.00 Matches: 248
 Percent Similarity: 57.04% Conservative: 84
 Best Local Similarity: 42.61% Mismatches: 184
 Query Match: 33.22% Gaps: 66
 DB: 6 Indels: 15
 US-09-856-327-2 (1-618) x ARI70485 (1-1701)
 QY 37 AspValPheIleAlaGlySerGlyProIleGlyAlaThrTyrAlaLysLeuCysValGlu 56
 Db 97 GATGTTTCATGCTGCGAGGTGCCATGATGCTGACTACTACGCCCGCACATCATTTGAC 156
 QY 57 -----AlaGlyLeuArgValValMetValGluIleGlyAlaAlaAspSerPheTyrAla 74
 Db 157 AATACCTCAACTACAAGGTTTACATGGCCAAATAGGTTCTCAAGATAC----- 207
 QY 75 ValAsnAlaGluGluGlyThrAlaValProTyrValProGlyTyrHisLysAsnGlu 94
 Db 208 -----CCTGTATC---GGGGCCCATCAACAACTCC 237
 QY 95 IleGluPheGlnLysAspIleAspArgPheValAsnValIleLysGlyAlaLeuGlnGln 114
 Db 238 ATAAGTTTTCAGAAAGCATTTGACAAAGTTTGTGAATATCATCAACAGGTGCCCTCCAGCG 297
 QY 115 ValSerValProValArgAsnGlnAsnValProThrLeuAspProGlyAlaThrSerAla 134
 Db 298 ATTTTGATTTTGGCCATGCGACCTACCAAGCCCACTCGCTGCTGAGCAGCGTGG---GCG 354
 QY 135 ProProGlySerSerAla-----IleSerAsnGlyLysAsnProHisGlnArg 150
 Db 355 CCGGCCATCATCTCTCCGGAAGCGCACCTGCTGATTATGGACACAAATCCGAATCAGAGAG 414
 QY 151 GluPheGluAsnLeuSerAlaGluAlaValThrArgGlyValGlyGlyMetSerThrHis 170
 Db 415 GCCGCGCTGAACCTTCCCGGTAGCGCTGTCACTGAGCAGTCCGGGGGATGGCAACCCAC 474
 QY 171 ThrThrCysSerThrProArgThrIleHisProMetGluSerLeuProGlyIleGlyArg 190
 Db 475 TGGACTTGGCGGTCTCTACTCCCATGAC----- 504
 QY 191 ProLysLeuSerAsnAspProAlaGluAspAspLysGluTrrPasnGluLeuTyrSerGlu 210

```

Db      505 GAGGAGGGGTCAACACCCAGTT---GACACAGCAGGAGTTGCAGCCATGCTGCAGCT 561
Oy      211 AlagluArgLeuilegLYThrSerThryLysGluPheAspGluSerIleIleArgThrIleu 230
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      562 GCTAAACACTTGTCTACAGCTTCACAGCCAGTCAGTACAGATTCATCCGTAGATAGTT 621
Oy      231 ValLeuArgSerLeuGlnAspAlaTyrLysAspArgGlnArgIlePheArgProLeuPro 250
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      622 GTCAAGAGACTCTTCAGACAGCCCTT---GATGCGTCGCGGGGTGTGACCACTGCCG 678
Oy      251 LeuAlaCysHisArgLeuLysAsnAlaProGluTyrValGluTrpHisSerIleGluAsn 270
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      679 CTGGGGGTGACGCCGTCAGACGACAATCTATTATGTACCTGACCGCTCCGAT--- 735
Oy      271 LeuPheHisSerIleTyrAsnAspAspLysGlnLysLysLeuPheThrLeuThrAsn 290
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      736 -----ACCGTCTTGCTGATGTGCCGAGAGACTCCCGGATTCGCTTGGTTACAGAG 786
Oy      291 HisArgCysThrArgLeuAlaLeuThrGlyTyrGluLysLysIleGlyAlaAlaGlu 310
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      787 ACGAGAGTACGACACTTATGTCTAGTGAACCAATCCGACGCGAGTTGTCTGCTGCTG 846
Oy      311 ValArgAsnLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIleMetAla 330
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      847 CTACGTAACCTG-----AATACACAGCAAC-----GATGAACTTGTCTGCGCC 888
Oy      331 LysValTyrValLeuAlaSerGlyAlaIleGlyAsnProGlnIleLeuTyrAsnSerGly 350
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      889 AAGACTTTCGTCATAGCTTGTGACAGACGTCGACACCGCAATCTGTGGAACAC--- 945
Oy      351 PheSerGlyLeuGlnValThrProArgAsnAspSerLeuIleProAsnLeuGlyArgTyr 370
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      946 -----ACATCGGCCCATAT-----GCGCTTGTGCTGCTAC 975
Oy      371 IleThrGlnGluProMetAlaPheCysGlnIleValLeuArgGlnGluPheValAspSer 390
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      976 CTCACGACGACGTCATGACTTTTGTGATGATGCTCTCAAGAGGGCATCTCATGATGCC 1035
Oy      391 ValArgAspAspProTyrGlyLeuProTrpTrpLysGluAlaValAlaGlnHisIleAla 410
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1036 ATCGCTAGTACGACCTGCGC-----TTCGCTGCGAAGTTGAGCGCGACAAAGAG 1083
Oy      411 LysAsnProThrAspAlaLeuProIleProPheArgAspProGluProGlnValThrThr 430
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1084 AAGCACCCGATGAGCTGCTGCCATTCATTCACAGACCCGAACTCAAGTCAATGAT 1143
Oy      431 ProPheThrGlnGlnHisProTrpHisThrGlnIleHisArgAspAlaPheSerTyrGly 450
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1144 CCGTACAGCTGCGACTTCCTTGCAATGTCAGGTGATCGGATGCAATTCATATGAT 1203
Oy      451 AlaValGlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThr 470
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1204 GATGTTGGACCCAGCGGACCGCGGTGTGTGTCGATGTCAGGTTTTCGCAATCA 1263
Oy      471 AspProGluAlaAsnAsnLeuValPhe-----Gln 481
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1264 GATATTGTGCGAAGAAATCGATGACTTTCGCTCCGAAACCTTAAGTCAGCGAGTGGGA 1323
Oy      482 AsnAspValGlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThr 501
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1324 GCGGGTGTTCACACACTTATGAAATGCCACAGCCGACATTCATGTCAGGAGCCAAAC 1383
Oy      502 AlaSerAsnValArgAlaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeu 521
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1384 GCCGATGAGACCGGTGACACAGATGATGATGATGATGATGATGATGATGATGATGAT 1443
Oy      522 GlyGlyTyrLeuProThrSerProProGlnPheMetAspProGlyLeuAlaLeuHisLeu 541
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1444 GGTGGGTACCTTCGCTGCTCAACCTCAATATTATGACGATGCTGCTGCTGCTGCTGCT 1503
Oy      542 AlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeu 561
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      1504 ACGGAACTACTCGGATCGGACAGATGATCAAACTTCTGTTCTGATCCGACATCAAG 1563
Oy      562 ValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGly 581
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1564 GTTCATTAACCTTCAACATGTGTGGGTGCGGGGAGATGGGTGATTCACAGATCGACTGCC 1623
Oy      582 GluAsnProThrLeuThrSerMetCysHisAlaIleLysSerAlaArgSerIleLeuAsn 601
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1624 TCCAAACCCGACTCGTACGAGCGCTGATGCGCTCAAGGGTGTGAGCTGTACTCAAT 1683
Oy      602 ThrLeu 603
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1684 TACCTT 1689
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
LOCUS      E33665
DEFINITION Novel protein, gene encoding the same and method for utilization
            thereof.
ACCESSION  E33665
VERSION    E33665.1 GI:18624149
KEYWORDS   JP 2000083675-A/1.
SOURCE     Tricholoma matsutake
            Tricholoma matsutake
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Tricholomataceae; Tricholoma.
REFERENCE  1 (bases 1 to 1946)
            Takakura,Y., Kuwata,S. and Ota,S.
            Novel protein, gene encoding the same and method for utilization
            Patent: JP 2000083675-A 1 28-MAR-2000;
            JAPAN TOBACCO INC. SOCIETY FOR TECHNO-INNOVATION OF AGRICULTURE
            FORESTRY AND FISHERIES
COMMENT     OS Tricholoma matsutake
            PN JP 2000083675-A/1
            PD 28-MAR-2000
            PF 08-SEP-1998 JP 1998270606
            PR YOSHIMITSU TAKAKURA, SHIGERU KUWATA, SYOZO OTA
            PT C12N15/09,A01N37/12,A01N63/00,A01N63/02,A01N65/00,C07K1/14, PC
            C07K1/18,
            PC C07K1/30,C07K14/375,C12N1/19,C12N1/21,C12N5/10,C12N5/10, PC
            C12P21/02,
            PC C12Q1/68//((C12N15/09,C12R1:645),C12N15/00,C12N5/00,C12N5/00,
            CC
            CC Key location/Qualifiers
            FH FT source 1..1946
            FT location/Qualifiers
            FEATURES
            source 1..1946
            location/Qualifiers
            /organism="Tricholoma matsutake"
            /mol_type="genomic DNA"
            /db_xref="taxon:40145"
BASE COUNT 484 a 516 c 475 g 471 t
ORIGIN
Alignment Scores:
Pred. No.: 4,13e-72 Length: 1946
Score: 1044.00 Matches: 241
Percent Similarity: 55.91% Conservative: 90
Best Local Similarity: 40.71% Mismatches: 187
Query Match: 31.79% Indels: 74
DB: 6 Gaps: 17
US-09-856-327-2 (1-618) x E33665 (1-1946)
Oy      27 IleHisGluThrTyrGlyAsnAspGlyValAspValPheIleAlaGlySerGlyProIle 46
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      185 GTACATTCACACT-----GATGTTTCATGCTGCGGAGTGTCCCAT 226
Oy      47 GlyAlaThrTyrAlaLysLeuGlu-----AlaGlyLeuArgValIleMetAl 64
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      227 GCTGTACTTACGCCCGCACATCAATGATCACTCAACAAAGGTCTACATGCGCC 286

```

QY	65	GlutIleGlyAlaAlaIleAspSerPheTyrAlaValAsnAlaGluGluGlyThrAlaValPro	84
Db	287	GAATAGGCTTCAGATTAAC-----	-CCT 310
QY	85	TyrValProGlyTyrHisLysAsnGluIleGluPheGlnLysAspIleAspArgPhe	104
Db	311	GTCAAC---GGAGCCATCACAGAACTCCATTAAGTTTCAGAAAGACCTACAAAGTTT	367
QY	105	ValAsnValIleLysGlyAlaLeuGlnGlnValSerValProValArgAsnGlnAsnVal	124
Db	368	GTGAATATCATCAACAGGTGCCCTCCAGCCCATTTGCATTTGCCATTCGGACACCTAACAG	427
QY	125	ProThrIleAspProGlyAlaTyrSerAlaProProGlySerSerAla-----	140
Db	428	CCCACTCTCGCTGTAGCAGCGGG---GCGCGGCCCATTCGATCTGCGGAAGCCAGCTC	484
QY	141	IleSerAsnGlyLysAsnProHisGlnArgGluPheGlnAsnLysSerAlaGluAlaVal	160
Db	485	GTGATATTAGGGACACAAATCCGAATGAGAGCGCGCCGACACTTCCCGGTAGCGCTGC	544
QY	161	ThrArgGlyValGlyGlyMetSerThrHisTyrPheCysSerThrProArgIleHisPro	180
Db	545	ACGAGGACAGTCGGGGGAATGGCGACCCACATGGAATTCGCGGTCTCACTCCACATGAC	604
QY	181	ProMetGluSerLeuProGlyTyrLeuTyrProLysLeuSerAsnAspProAlaGluAsp	200
Db	605	-----GAGAGAGGGTCAACACCCACTT---GAC 631	
QY	201	AspLysGluTyrPAsnGluLeuTyrSerGlnAlaGluArgLeuIleGlyThrSerThrLys	220
Db	632	AAGCGAGGTTCGACGACGACTGTCGAGCGTGCTAAACATTCCTCAAGTTCACAGCGAC	691
QY	221	GluPheAspGluSerIleArgHisThrLeuValLeuArgSerLeuGlnAspAlaTyrLys	240
Db	692	CAGTGTGACGATCTCATCCGTGACATGTGTCAAAAGAACCCCTTCACACACCTT--- 748	
QY	241	AspArgGlnArgIlePheArgProLeuProLeuAlaCysHisArgLysAsnAlaPro	260
Db	749	GATGCGTCGGGGGTGACCACTCTCCGGTGGGGGGGAGCGCGGACGACAAATCT	808
QY	261	GluTyrValAlaGluTyrHisSerAlaGluAsnLeuPheHisSerIleTyrAsnAspLys	280
Db	809	ATTATATGACCTGACCGCGCCGAT-----ACCGTCTTGATGATGGCCG 856	
QY	281	GlnLysLysLeuPheThrLeuLeuThrAsnHisArgCysThrArgLeuAlaLeuThrGly	300
Db	857	AAGAATCCCCCATTCGTTGGTTTACAGACGAGAGAGACGAATTTATTTGTCACTGAA	916
QY	301	GlyTyrGluLysLysIleGlyAlaAlaGluValArgAsnLeuValThrArgAsnPro	320
Db	917	ACCAATCCGACGACAGGTGTCGTCGTTGCTACTACTACTG-----AATACA 964	
QY	321	SerSerGlnLeuAspSerTyrIleMetAlaLysValLysValLeuAlaSerGlyAlaIle	340
Db	965	AGCAAC-----GATGAACCTGCTGGGCCCAAGAACTTCGTCATGCTGTGGACACGT	1018
QY	341	GlyAsnProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProArgAsn	360
Db	1019	TGCACACCGCAAAATCTGTGGAAAGC-----AACATCCGCCCAAT--- 1066	
QY	361	AspSerLeuIleProHisLeuGlyArgTyrIleThrGluGlnProMetAlaPheCysGln	380
Db	1061	-----GCGCTTGCTGCTCACTCCACGAAACAGTCATGACTTTTGTGTGAG 1105	
QY	381	IleValLeuArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuProThr	400
Db	1106	ATTGTGTCCAAGAGAGCATAGTCATTCATTCGCTACTGACCTCGC----- 1155	
QY	401	TrpLysGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeuProIlePro	420
Db	1154	TTGCGTCGGAAGTTGAGCGCACAGAAGAACCCCGATGACGTCTGCGCGATTTCA 1211	

OY		421	thearydarpProgluProglInvalThrThrProhetrhruGlunhsProtpRhIsThr	440
Db		1214	TTCCAGCGACGCTGGAACTCAAGTATGAATTCGGACCTCCCTGGCATGT	1273
OY		441	GlnIleHtSaAgAsPaLaPheSerTyrcLyAlaValGlyProgluValAspSerArgVal	460
Db		1274	CAGGTCCATCCG-----TATGCATTGGTGATGTTGGACCCAAAGGCCGACCCGGCTTT	1327
OY		461	IleValAspLeuArgTrpPheGlyAlaThrAspProgluAlaAsnAsnLeuValPhe	480
Db		1328	GTCTGCATGTAGAGTTTTCGGCAAAATCAGATATTGCGAAGAANAATCGATGACTTTC	1387
OY		481	-----GlnAsnAspValGlnAspGlyTyrSermEpro	491
Db		1388	GGTCGGAACCCTAAGCTACGGCACTGGGAGACGGGGTGTACAGACACTTATGGAATGCCA	1447
OY		492	GlnProthrPheArgTrpArgProSerThrLaseAsnValAcgAlaArgLysMetet	511
Db		1448	CAGCCGACATTTCATGTTCACGGGACCAGCCCGATGAGACCGGACCGACAGAGATGATG	1507
OY		512	AlaAspMetCySGluValAlaSerAsnLeuGlyGlyTyrLeuProthrSerProProglIn	531
Db		1508	AATGATATAGCACCAACGTCGCAACATACTGGGCGGATCTCTCGCTCTACCTCA	1567
OY		532	PheMetaspproGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLys	551
Db		1568	TTTTATGGACACCTGGCTCGCACACACATCCAGGAACTACTCGATGGGACAGATGAT	1627
OY		552	AlaThrThrValAlaAspAsnAsnSerLeuValTrpaspPheHlaAsnLeuTyrValAla	571
Db		1628	CAACCTTCGTGGTCGATCCGACATCAAAGGTTCTATAACTCGACAATCTGGGTGGCGC	1687
OY		572	GlyAsnGlyThrIleArgThrglyGlyLysAsnProThrLeuThrSermCysHis	591
Db		1688	GGGAATGGGTGCATTCGATCCAGATGCCAGCTCCTCGCAACCGCACTCGACAGCGTCCGAT	1747
OY		592	AlaIleLysSerAlaArgSerIleLeuAsnThrLeu	603
Db		1748	CGCGTTAAGGTGCTGACGCTGTAGTCACTTACCTT	1783
RESULT 11				
E37379				
LOCUS	E37379	1946 bp	DNA	linear
DEFINITION	Reagent for assaying pyranose containing pyranose oxidase.			PAT 31-JAN-2002
ACCESSION	E37379			
VERSION	E37379.1 GI:18626680			
KEYWORDS	JP 2000175698-A/1.			
SOURCE	Tricholoma matsutake			
ORGANISM	Tricholoma matsutake			
	Eukaryota; Fungi; Basidiomycota; Hymenomyces; Homobasidiomycetes;			
	Agaricales; Tricholomataceae; Tricholoma.			
REFERENCE	1 (bases 1 to 1946)			
AUTHORS	Takakura,Y., Kuwata,S., Ota,S. and Usami,S.			
TITLE	Reagent for assaying pyranose containing pyranose oxidase			
JOURNAL	Patent: JP 2000175698-A 1 27-JUN-2000;			
	JAPAN TOBACO INC			
COMMENT	OS Tricholoma matsutake (pine mushroom)			
	PN JP 2000175698-A/1			
	PD 27-JUN-2000			
	PF 16-DEC-1998 JP 1998357423			
	PR			
	PI YOSHIMITSU TAKAKURA,SHIGERU KUWATA,SHOZO OTA,SATORU USAMI PC			
	C1201/26,C12N9/04,C12N15/09/G01N33/66,(C12N9/04,C12R1:645),PC			
	(C12N15/09,C12R1:645),C12N15/00,(C12N15/00,C12R1:645) CC			
	FH Key			
	Location/Qualifiers			
FT	1..1946			
FT	source	/organism="Tricholoma matsutake (pine		
	FT	mushroom)"		
		location/Qualifiers		
FEATURES		1..1946		
SOURCE		/organism="Tricholoma matsutake"		
		/mol_type="genomic DNA"		
		/db_xref="taxon:40145"		

TITLE	Purification, characterization and primary structure of a pyranose
JOURNAL	oxidase from <i>Tricholoma matsutake</i>
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 1946)
TITLE	Takakura, Y.
JOURNAL	Direct Submission
	Submitted (26-MAY-2000) Yoshimitsu Takakura, JT Plant Innovation
	Center, 700 Higashibara Toyoda-cho, Iwata-gun, Shizuoka-ken
	438-0800, Japan (E-mail: yoshimitsu.takakura@ims.jti.co.jp,
	Tel:81-538-32-8291, Fax:81-538-33-6046)

BASE COUNT	484 a	516 c	475 g	471 t
ORIGIN				

Alignment Scores:	
Pred. No.:	4.13e-72
Score:	1044.00
Percent Similarity:	55.91%
Best Local Similarity:	40.71%
Query Match:	31.79%
BB:	8
Length:	1946
Matches:	241
Conservative:	90
Mismatches:	187
Indels:	74
Gaps:	17

US-09-856-327-2 (1-618) x AB043883 (1-1946)

QY	27	IIHISGLIHTHTyTcGlyAsnAspGlyAlaPheIleAlaGlySerGlyProIle	46
Db	185	GRACATTACACT-----GATGTTTCATTGGTCAGAGGCGCCATT	226
QY	47	GIYAlAThTyrAlaLysLeucGysValGlu-----AlaGlyLeuArgValIleMetVal	64
Db	227	GCCTGACTAAGCGCCGCGACATCATTTGCACAACTACCTACAAAGCTCFACATGGCC	286
QY	65	GIuIIeGlyAlaAlaAspSerPheTyrAlaValAsnAlaGluGluGlyThrAlaValPro	84
Db	287	GAATAGGTTCTCCAGATAC-----CCT	310
QY	85	TyrValProGlyTyrHisLysAsnGluIleGluPheGlnLysAspIleAspArgPhe	104
Db	311	GTGTCAT--GGAGCCCATCCAGAAATCCCTAAAGTTTCGAAAGACACGTCAACAGTTT	367
QY	105	ValAsnValIleLysGlyAlaLeuGlnGlnValSerValProValArgAsnGlnAsnVal	124
Db	368	GRGAATATATATCAAGGGRCCCTCCAGCCCAATTCGATTTGGCATGGAGACCTACACAG	427
QY	125	ProThrLeuAspProGlyAlaATrPheSerAlaPropGlySerSerAla-----	140
Db	428	CCACACTCTGGCTGAGACAGCTGG---GCGCGCCCATCATGCTCTGCCGAAGGCGACACTC	484
QY	141	IleSerAsnGlyLysAsnProHisGlnIleArgGluPheGluAsnLeuSerAlaGluAlaVal	160

[illegible]

Db 1448 CACCCACATTCATGCAAGCGGACCAACCCGATGAGACCGTGACCAAGAGATGATG 1507
 QY 512 A1aSpMeTcysGluValAlaASerAsnLeuGlyIYrLeuProThrSerProProGln 531
 Db 1508 AATGATATGACCAACGCGGACATATGCGGGGAGTACTCTGCTGCTCTCAACCTCA 1567
 QY 532 PhemeTAspProClyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLys 551
 Db 1568 TTTATGGCACCTGCTCTCGACAGCATACGGAACATCTCGATCGGACGACATGAT 1627
 QY 552 AlArThrThyValAlaAspAsnSerLeuValITrpaSppHleaLsnLeuTyValAla 571
 Db 1628 CAACACTTCTGTCTGATCCGACATCAAGTTCAATCTTCACATCTCTGTGGTCCG 1687
 QY 572 GlyAsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHis 591
 Db 1688 GGGAAATGGTGATCCAGATGCGACCTGCTGCAACCGCGCTGATGACGACGTCGCTAT 1747
 QY 592 AlAlaIleYserAlaArgSerIleIleAsnThrLeu 603
 Db 1748 GCGCTTAAGGCTGCTGAGCTGATGATGCTTACCTT 1783

RESULT 13
 ANI507836 165137 bp DNA linear BCT 12-MAR-2003
 ANI507836
 DEFINITION Arthrobacter nicotinovorans PA01 plasmid sequence.
 ACCESSION AJ507836
 VERSION 1
 KEYWORDS 2,6-dihydroxypyridine hydroxylase; 2-keto-gluconate dehydrogenase-like protein; 6-HNO gene; 6-HNO gene; 6-hydroxy-D-nicotine oxidase; 6-hydroxy-L-nicotine oxidase; amine oxidase; amino acid permease; amino acid transporter; ATP binding component of ABC transporter; ATP-binding protein; ATP/GTP-binding protein; beta-glucosidase; carbon monoxide dehydrogenase subunit D; carbon monoxide dehydrogenase subunit E; carbon monoxide dehydrogenase subunit F; carbon monoxide dehydrogenase subunit G; CGL gene; chromosome partitioning protein; conjugative transfer gene complex protein; cox2 gene; cox3 gene; cox4 gene; cox5 gene; CSCA gene; DnaH3 gene; dhph gene; DNA helicase-like protein; DNA ligase; DNA recombinase; ebrA gene; ebrB gene; endopeptidase; ESTa gene; formyltetrahydrofolate deformylase; glycine cleavage system T protein; hydroxylase; inositol transport protein; integral membrane protein; Kds gene; KDH gene; KDH1 gene; KDH2 gene; KDH3 gene; ketone dehydrogenase (large subunit); ketone dehydrogenase (medium subunit); ketone dehydrogenase (small subunit); KtrA-like protein; LacI-family transcriptional regulator; LacI-family transcriptional regulatory protein; lipase (esterase); methyltetrahydrofolate cyclohydrolase; methyltetrahydrofolate dehydrogenase; MUR gene; MOA gene; MOA gene; MOA gene; MOA gene; MOA gene; MODC gene; MODC gene; molybdate-binding periplasmic protein modA; molybdenum transport ATPase modC; molybdenum transport transmembrane protein modB; molybdopterin cofactor synthesis protein moaA; molybdopterin cofactor synthesis protein moaB; molybdopterin cofactor synthesis protein moaC; molybdopterin cofactor synthesis protein moaD; molybdopterin cofactor synthesis protein moaE; molybdopterin synthase (large subunit moaB); multidrug resistance efflux protein; NAD-dependent aldehyde dehydrogenase; NADPH:quinone oxidoreductase; NDH gene; NDHM gene; NDH gene; nicotine dehydrogenase (large subunit); nicotine dehydrogenase (medium subunit); nicotine dehydrogenase (small subunit); nitrile amino hydrolase; Nrh gene; OR0494 gene; ORF101; ORF103; ORF105; ORF106; ORF107; ORF110; ORF114; ORF116; ORF117; ORF119; ORF133; ORF137; ORF145; ORF146; ORF147; ORF153; ORF155; ORF16; ORF164; ORF165; ORF167; ORF168; ORF17; ORF170; ORF175; ORF176; ORF18; ORF180; ORF190; ORF192; ORF194; ORF196; ORF201; ORF203; ORF204; ORF208; ORF21; ORF216; ORF218; ORF22; ORF221; ORF223; ORF224; ORF226; ORF231; ORF232; ORF236; ORF24; ORF246; ORF249; ORF25; ORF26; ORF264; ORF27; ORF287; ORF29; ORF294; ORF30; ORF300; ORF302; ORF308; ORF309; ORF31; ORF319; ORF311; ORF314; ORF32; ORF323; ORF327; ORF33; ORF332; ORF339; ORF350; ORF357; ORF36; ORF367; ORF37; ORF38; ORF385; ORF388; ORF389; ORF395; ORF394; ORF406; ORF407; ORF421; ORF424; ORF436; ORF458; ORF465; ORF470; ORF473; ORF48; ORF480; ORF481; ORF487; ORF497; ORF502; ORF530; ORF549; ORF583;

SOURCE ORF594; ORF615; ORF64; ORF653; ORF666; ORF67; ORF68; ORF70; ORF72; ORF78; ORF82; ORF85; ORF87; ORF89; ORF890; ORF93; ORF96; ORF97; ORF99; oxidoreductase; PAR gene; permease; polyketide cyclase; PURU gene; RBSA gene; recombinase; redoxin; ribose ABC transporter; RNA polymerase sigma factor; single-stranded DNA-binding protein; smf family protein; soj family protein; SSB gene; succinate-semialdehyde dehydrogenase; sucrose hydrolase; sugar ABC transporter; YED gene; transcriptional regulator; transfer gene complex protein-like protein; transferase; transporter; transposase; WZT gene.

ORGANISM Arthrobacter nicotinovorans

REFERENCE 1 Micrococciinae; Micrococaceae; Arthrobacter.

AUTHORS Grether-Beck, S., Igloi, G.L., Pust, S., Schilz, E., Decker, K. and Brandsch, R.

TITLE Structural analysis and molybdenum-dependent expression of the PA01-encoded nicotine dehydrogenase genes of Arthrobacter nicotinovorans

JOURNAL Mol. Microbiol. 13 (5), 929-936 (1994)

MEDLINE 95115562

PUBMED 7815950

REFERENCE 2 (bases 1 to 165137)

AUTHORS Menendez, C., Igloi, G., Hemminger, H. and Brandsch, R.

TITLE A PA01-encoded molybdopterin cofactor gene (moa) of Arthrobacter nicotinovorans: characterization and site-directed mutagenesis of the encoded protein

JOURNAL Arch. Microbiol. 164 (2), 142-151 (1995)

MEDLINE 96172783

PUBMED 8586735

REFERENCE 3 (bases 1 to 165137)

AUTHORS Menendez, C., Igloi, G.L. and Brandsch, R.

TITLE IS1473, a putative insertion sequence identified in the plasmid PA01 from Arthrobacter nicotinovorans: isolation, characterization, and distribution among Arthrobacter species

JOURNAL Plasmid 37 (1), 35-41 (1997)

MEDLINE 97230479

PUBMED 9073580

REFERENCE 4 (bases 1 to 165137)

AUTHORS Menendez, C., Otto, A., Igloi, G., Nick, P., Brandsch, R., Schubach, B., Botcher, B. and Brandsch, R.

TITLE Molybdate-uptake genes and molybdopterin-biosynthesis genes on a bacterial plasmid--characterization of MoaA as a filament-forming protein with adenosinetriphosphatase activity

JOURNAL Eur. J. Biochem. 250 (2), 524-531 (1997)

MEDLINE 98088982

PUBMED 9428706

REFERENCE 5 (bases 1 to 165137)

AUTHORS Schenk, S., Hoeltz, A., Krauss, B. and Decker, K.

TITLE Gene structures and properties of enzymes of the plasmid-encoded nicotine catabolism of Arthrobacter nicotinovorans

JOURNAL J. Mol. Biol. 284 (5), 1323-1339 (1998)

MEDLINE 99068870

PUBMED 9878353

REFERENCE 6 (bases 1 to 165137)

AUTHORS Baisch, D., Sandu, C., Brandsch, R. and Igloi, G.L.

TITLE Gene cluster on PA01 of Arthrobacter nicotinovorans involved in degradation of the plant alkaloid nicotine: cloning, purification, and characterization of 2,6-dihydroxypyridine 3-hydroxylase

JOURNAL J. Bacteriol. 183 (18), 5262-5267 (2001)

MEDLINE 21405725

PUBMED 11514508

REFERENCE 7 (bases 1 to 165137)

AUTHORS Igloi, G.L. and Brandsch, R.

TITLE Sequence of the 165-kilobase Catabolic Plasmid PA01 from Arthrobacter nicotinovorans and identification of a PA01-dependent Nicotine Uptake System

JOURNAL J. Bacteriol. 185 (6), 1976-1986 (2003)

MEDLINE 22505657

PUBMED 12618462

REFERENCE 8 (bases 1 to 165137)

AUTHORS Igloi, G.L. and Brandsch, R.

TITLE
JOURNAL

Direct Submission
Submitted (12-SEP-2002) Institute of Biochemistry and Molecular
Biology, University of Freiburg, Hermann-Heider-Str. 7, Freiburg
D-79104,
GERMANY

FEATURES

Location/Qualifiers
1. 165137

source

/organism="Arthrobacter nicotinovorans"

/mol_type="genomic DNA"

/db_xref="taxon:29320"

/plasmid="pAO1"

/complement(1. .296)

/note="ORF13"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="hypothetical protein"

/protein_id="CAD47859.1"

/db_xref="GI:25169023"

/translation="MIRRTATKRRRORAGPQTAELCDGTGGRNRPKRGVDAFPG
SPGPPPTSPSTSPGEGDGRPHLPPEGLVRAAVRRRQPPAPRRRTAGS"

/complement(344. .1138)

/note="ORF264"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="putative hydrolase"

/protein_id="CAD47860.1"

/db_xref="GI:25169024"

/translation="MPTPTSGIYWGSGPVAAPAVLLIEGYTGQILGRDAFCDELLA
QGLRYLRMDNRDVGLSREDGNMYADMAVDIVADADIGKITIVQSGMGLIAQ
TALGYPMNVGLIVETPTVLDIDPGILITAEIPPOREDAIASFLEDGRATAPAN
GYDEAKRELGLVMEFRAPDRSGLSRQNAVALPDLPRRLTELMFVALLHGRNDAL
IRAGSLRIAEVPOAELHLYPGMGHEIAIPALMDEFVAITIRIAC"

/complement(1140. .2093)

/gene="CSCA"

/complement(1140. .2093)

/gene="CSCA"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="putative sucrose hydrolase"

/protein_id="CAD47861.1"

/db_xref="GI:25169025"

/translation="MLAPEDRWIMDFMHVODHGIIHLYLQAPKSLDPPELRRNATI
GHATSTNLDMWTEGTVLRPGGCGTVDAATWPGSVVAGSDGDMPEMTGSAFLSPDS
ATNVETIVAVASDILHHAKADPTALADSAWTEKLGSSPPEERAMPWYRTDTGL
WHMLTARANTGEVNERGVGHATSQDILRAMTARPLPGEFHNLEVIQSVINDQ
TFMIFSAHETVVAQRCDRAHATGMAVWDRIELEHARNITGPTELYSGRIAHNGL
PVLLGFLLFSGGGITGSDPVDLEGLLKSGLLSPQAG"

/complement(2124. .3092)

/gene="ESTa"

/complement(2124. .3092)

/gene="ESTa"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="putative lipase (esterase)"

/protein_id="CAD47862.1"

/db_xref="GI:25169026"

/translation="MPVLPALQTLDSLPATGLDPTDLPVAVRQAGLEGANALALVA
OEPIDVGSVHTAMVNGEIKVRIYQGTGPHPIHVFYHGGGMAGTIEETFDVNG
CREETALAGYTVAVAEYRLAPEFRPILEDSYALAEVVEHADYDGPENVTGGG
SAGGNLSAAILAKRANENGPHTROYLEVPALDITLAPSVYSGGGYLPKPAITDL
CVEGYLRSPEDARNPYASPLADVSGPLPALITISSEDPLOMDSRYAQRLOEYGVH
AAFRLEGHVGSSQFTKLPEPAAARQVLDALKTHADHRRTVP"

/complement(3148. .4482)

/gene="OB0494"

/complement(3148. .4482)

/gene="OB0494"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="putative inositol transport protein"

/protein_id="CAD47863.1"
/db_xref="GI:25169027"
/translation="MSTQIDPETGSSPAGAAAPVAKYHPRVAVIAGLASYVGCAL
TVNGIALVITQOTIGLTDYGLLTAMTVGIALGALGKRGDILGKRFATIMAV
IMLSLAPFTSDFTVLLAGITLLGLGVADLPVSLATIAETKSRKMLVFYQVM

Alignment Scores:

Score:	8.18e-07	Length:	165137
Percent Similarity:	243.00	Matches:	159
Best Local Similarity:	35.14%	Conservative:	68
Query Match:	24.61%	Mismatches:	228
DB:	7.40%	Indels:	191
	1	Gaps:	28

US-09-856-327-2 (1-618) x ANT507836 (1-165137)

QY	10	ArgAspTYrProAlaGSerMetGlnLeaSnGlyGlnIleProLYsAsnAlaIleHisGlu	29
DB	34222	CGACCTCTCCCAACCATGCTCTCGACGTGACCGAATATGCAGACAGCCGATGAAA	34281
QY	30	ThrTYrGly-----AsnAspGlyValAspValPheIleAlaGlySerGlyProIleGly	47
DB	34282	AAATATGACCTCCTCATGATGACGACGCTCCCTTATCATC--GGTCCGGCCCGCGCGT	34338
QY	48	AlaThrTYrAlaLYsLeuCYsValGluAlaGlyLeuArGValMetValGluIleGly	67
DB	34339	GCCACCTTGCCCTTCAGACTGCTGCAGCAAGGCTTAAAGTGTCTTTCGACACAGGC	34398
QY	68	AlaAlaAspSerPheTYrAlaValAsnAlaGluGluGlyThrAlaValProTYrValPro	87
DB	34399	-----CGATGATGCACC	34410
QY	88	GlyTYrHisLYsLYsAsnGluIleGluPheGlnLYsAspIleAspArgPheValAsnVal	107
DB	34411	AAC-----GACAGTTCATCAACGAC	34431
QY	108	IleLYsGlyAlaLeuGlnGlnValSerValProValArgSnGlnAsnValProThrIleu	127
DB	34432	GAACGGACGCTCTTACACAGCTCA-----TGCAGCGACACCCGACTG	34476
QY	128	AspProGlyAlaTYrSerAlaPro-----ProGlySerSerAlaIleSerAsnGly	144
DB	34477	GCACTGGATCATGAGCCTCGCGAAGCACTTCCCGGACGCCGCGC---TGAAACGGA	34533
QY	145	LYsAsnProHisGlnAlaGluPheGluAsnLeuSerAlaGluAlaValThrArgLYsAl	164
DB	34534	AAG-----GCCGTC	34542
QY	165	GlyGlyMetSerThrHisTYrThrCYsSerThrProArgIle-----	178
DB	34543	GGTGGACGGCGACATTTCTGGACCGGACTGACACCCCATTTCAATGGCAGACTTCAAG	34602
QY	179	---HisProPromeGluSerIleProGlyIleGlyArgProLYsLeuSerAsnAspPro	197
DB	34603	ACACACAGCTACTACGAGACCTGCCCAT-----GGAACATGCGACAGCTGGCCA	34653
QY	198	AlaGluAspAspLYsGluTYrPAsnGluLeuTYrSerGluAlaGluArgLeuIleLYsThr	217
DB	34654	CTGGACCTGGAC---GAACCTGCACCATTAACACGGCGAAAGAGCGGTGGAGCC	34710
QY	218	Ser-THrLYsGluPheAspGluSerIleArgHisThrLeuValLeuArg-SerLeuGlnA	237
DB	34711	TCCACCGGACGCGGACCTCCGTCGGCGGACCAAGGCTACAAAGTTCT-----	34763
QY	237	sPAlaTYrLYsAspArgGlnArgIlePheArgProLeuProLeuAlaCYsHisArgLeuL	257
DB	34764	-----GGCGAAGCGCGCGAAGCAATGGATACCGCATTAACGACGCGCTAC	34815
QY	257	ysAsnAlaProGluTYrValGluTYr-----HisSerAlaGluAsnLeuPheHisS	274
DB	34816	GCCACCAAGCTTGAACCTCAGTGGCGCGCGGCGGACAGTTCAGACGAGCTT-----	34868
QY	274	erIleTYrAsn-AspAspLYsGlnLYsLYsLeuPheThrLeuLeuThrAsnHisArgCYs	293

Db 34869 --CGTATGGCAGCGGATTAAGCCGCCAGTGGACACACTTGCACGAGATACCA 34926
 QY 294 ThrArgLeuAlaLeuThrGlyGlyTyrGluLysIleGlyAlaIleGlyAlaArgAsn 313
 Db 34927 AAGGCTTACG---ACGGCCCTGCTGGAACCTGGGACCGAACCCGACGCTACAA--- 34980
 QY 314 LeuLeuAlaThrArgAsnProSerSerGlnLeuAspSerTyrIle----- 328
 Db 34981 -----ATCACGCTGGCCCTGACGGCGCGCGCTGCTACACAGACACCGCTC 35034
 QY 329 -----MetAlaLysValTyrValLeuAlaSerGlyAlaIleGlyAsn 342
 Db 35035 GGAAACCTGCACAGCTAACCGGCTCGCTGTAGCTGTACAGAAACGCAATCGAGACC 35094
 QY 343 ProGlnIleLeuTyrAsnSerGlyPheSerGlyLeuGlnValThrProAlaGlnAspSer 362
 Db 35095 CCACGGCTGCTGCTGCTTGGCTTGGCTACCTCGGGC-----CATACCGACGCC 35139
 QY 363 Leu----- 363
 Db 35140 CTTCGCCAACTCTCAGGCGCAGTCCGGGCGCACTATATCGCGCACACACCGGCTCTGTC 35199
 QY 364 -----IleProAsnLeuGlyArgTyrIleThrGlnIleProMetAlaPheCysGln 380
 Db 35200 TACGCCAGTTCCTCCCAACGAGTCCATATGTACCGCGGCGCAACCAATGACGAGCATC--- 35256
 QY 381 IleValLeuArgGlnGlnIlePheValAspSerValArgAspAspPro-----TyrGly 397
 Db 35257 -----ATCAGCGACGATCGCGCATGCGGATGACCCATCAGCGGCTCTGCTT 35298
 QY 398 LeuProTTrpTTrpGluAlaValAlaGlnHisIleAlaLysAsnProThrAspAlaLeu 417
 Db 35299 GGGCGTACTACATCCAGATGATCGCCCAAGGCGCTGCCCTCATCTAGTAC----- 35349
 QY 418 ProIleProPheArgAspProGlnValThrProPheThrGlnGlnHisPro 437
 Db 35350 -----TTCATGAGCCCG----- 35361
 QY 438 TrpHisThrGlnIleHisArgAspAlaPheSerTyrGlyAlaValGlyProGluValAsp 457
 Db 35362 -----GGGAGTGGGGGCGCACAAATTCACA 35385
 QY 458 SerArgValIleValAspLeuArg-----TrpPheGlyAlaThrAsp---Pro 472
 Db 35386 GAAAGGTGAGAACCTTACACCCCGCACAGCGGCACTTGGATTTCGCGTGAGAGCTTCCG 35445
 QY 473 GluAlaAsnAsnLeuLeuValPheGlnAsnAspValGlnAspGlyTyrSerMetProGln 492
 Db 35446 CAGGCATCCACAGAGATGACCCCTAACGACACCGCTACCGACCGCTCGTCCGCCA 35505
 QY 493 ProThrPheArgTyr---ArgProSer-----ThrAlaSerAsnValArg 507
 Db 35506 CCTGTGCTCCACTACGACGACATCCCAACGACATCGCCATGCGGAACACGCGCTACGAG 35565
 QY 507 IaaArgLysMetMetAlaAspMetCysGluValAlaSerAsnLeuGlyTyrLeuProT 527
 Db 35566 CAGGCGGAGAACTCTCAATCTGTCG----- 35594
 QY 527 HisSerProProGlnPheMetAspProGlyLeuAlaLeu---HisLeuAlaGlyThrThr 545
 Db 35595 --AGCATCCGCGCAACACAGAGCCCGGGGATGCCCTCCGGCCACAACTGGGAAACGCC 35652
 QY 546 ArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsnSerLeuValTrpAspPhe 565
 Db 35653 CGAATGAGCGAAGGCTGATGAAGGCGTGTCAACAGTTTTCGCTCAAGCCGACGACGTA 35712
 QY 566 AlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgThrGlyPheGlyGlnAsnProThr 585
 Db 35713 CCTAACCTGTCTTCGACGAGGTCACAATTCACACGCGGCTCGCGGAACCCACCC 35772
 QY 586 LeuThr 587
 |||||

Db 35773 CTGACA 35778
 RESULT 14
 AE016780
 LOCUS
 DEFINITION
 Pseudomonas putida KT2440 section 7 of 21 of the complete genome.
 ACCESSION
 AE016780 AE015451
 VERSION
 AE016780.1 GI:26557024
 KEYWORDS
 Pseudomonas putida KT2440
 Pseudomonas putida KT2440
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 REFERENCE
 AUTHORS
 Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
 Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
 Beaman,M., Deboy,R., Daugherty,S., Kolonay,J., Madupu,R.,
 Nelson,M., White,O., Peterson,J., Khouiri,H., Hance,I., Lee,P.,
 Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Ullrichback,T.,
 Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
 Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
 Duesterhoft,A., Tumbler,B. and Fraser,C.
 Complete genome sequence and comparative analysis of the
 metabolically versatile Pseudomonas putida KT2440
 Environ. Microbiol. 4 (12), 799-808 (2002)
 2 (bases 1 to 305520)
 Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
 Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
 Beaman,M., Deboy,R., Daugherty,S., Kolonay,J., Madupu,R.,
 Nelson,M., White,O., Peterson,J., Khouiri,H., Hance,I., Lee,P.,
 Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Ullrichback,T.,
 Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
 Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
 Duesterhoft,A., Tumbler,B. and Fraser,C.
 Direct Submission
 Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1. 305520
 /organism="Pseudomonas putida KT2440"
 /mol_type="genomic DNA"
 /strain="KT2440"
 /db_xref="taxon:160488"
 108. .1397
 /gene="eno"
 /locus_tag="Pp1612"
 108. .1397
 /gene="eno"
 /locus_tag="Pp1612"
 /note="similar to SP:P08324; identified by sequence
 similarity; putative"
 /codon_start=1
 /trans_table=11
 /product="enolase"
 /protein_id="AAN67233.1"
 /db_xref="GI:24983094"
 /translation="MAKIVDIRGREYLDISRGNPVEADVLLDNGIIGSACAPSGASTG
 SEALELRDGDKRSRYLKGVLKAVANINIPIDLGKDPSPDOKALDRMILIDSTEN
 KAKLGIANILAVSLAAAKAADODDPLFAHANLNGTGGQYSPMPKMIINGEHA
 DNNDVIOEPFQVPVGAKTFSDGIRMGTEIFHNILKAVLKARGINTVAVDGGGAPNINLAS
 NEDACATINAEYERKAGYKIGTPTTLALDCAAEFEYEDKYNISGGEKSPDAGCPADYL
 KGITFERPILITIEDGLDESDMGWMKLTIDKIGEVOLVGDLLFTNTIKLKGIEKGI
 GNSILIKFMQIGSLTETLEAIOMAKAAGTAAVISHRSGETEDSTIADLAIVGPAAGQIK
 TCSLRSRDSRYNQILRIEEDLGAKAAYRGRAEFRG"
 1566. .1847
 /locus_tag="Pp1613"
 1566. .1847
 /locus_tag="Pp1613"
 /note="similar to GP:9949793; identified by sequence
 similarity; putative"
 /codon_start=1
 /trans_table=11
 /product="conserved hypothetical protein"

gene
/protein_id="AA067234.1"
/db_xref="GI:24983095"
/translation="MRSPYMLFLVLLILGLIOYRLWNGSLAQTLEKQIADQNA
ENERLLERNRYLDAEVLLEIKKMETVEERARHHELGWKEGTELLPOLRQK"
1844..2551
/gene="ispd"
/locus_tag="pp1614"
1844..2551
/gene="ispd"
/locus_tag="pp1614"
/note="identified by match to TIGR protein family HMM
TIGR00453"
/codon_start=1
/transl_table=1
/product="4-diphosphocytidylyl-2C-methyl-D-erythritol
synthase"
/protein_id="AA067235.1"
/db_xref="GI:24983096"
/translation="MIDILPAFMAVIPAAGVAGARMAADRPKOYLELAGOULLESILDC
FLGHPALKGVVSTAEDDPYWGLRCASDPRIHCAAGGERADSVLNAALVLAHOGAA
DSDMVLYHDAAPMLARSDDLKLSLADDPVGLLAVPADTLKKRADSNGVSATVD
RSTIQAATYTPQMFLGALHRALECLVSDVYVDEASIHMSGQAPRLVEGRSDNIKY
TRPDLEMLRQRMWGRK"
complement(2679..3575)
/locus_tag="pp1615"
complement(2679..3575)
/locus_tag="pp1615"
/note="identified by match to PFAM protein family HMM
PF03466"
/codon_start=1
/transl_table=1
/product="transcriptional regulator, lysr family"
/protein_id="AA067236.1"
/db_xref="GI:24983097"
/translation="MSSRWEGIDEFVAVESGQPTAAARLGVSSSHISQIARLEDR
LQTRILYSTRRYTLSEAGQFLQHCRLQDGRREALRAMGLASEPKGLRMTCAVA
YGERFIPLVTRFMALYPOLRVDVLELNHTLDLHEGMDLIRLGRSLVATRIIA
PRMYLVCSPAYLEIRGPHSLSELARNCLVSGALQDQRELSOROGWNRCS
SGQVDAALOGTIGCOLPDYVLEHNLNSGALVSLLEAHQPNPAVVALTYQQRHLSF
KVRRLVYIKKEGLAGLPREYRFG"
3680..4795
/locus_tag="pp1616"
3680..4795
/locus_tag="pp1616"
/note="identified by match to TIGR protein family HMM
TIGR01751"
/codon_start=1
/transl_table=1
/product="D-isomer specific 2-hydroxyacid dehydrogenase
family protein"
/protein_id="AA067237.1"
/db_xref="GI:24983098"
/translation="WMKSRRAVAFAEAKKPLEIVEVDVAMFKAGEVLLRVASGVCHT
DAYTLGSADEPGIIPSLIHGGAIVAVGEGVSLAVGDHVIPLYTECGKCFCS
GKTMICQAIRATQGLMDGTFRSYGQOLFMYMTSTFSEYTLPEISVAKIOKE
APLEKVCILGSGVTTIGAVLNTAKVKGPDVAFEGGIGLSAVIGAVKAKRIIA
IDINPAKFETAIROLGATDVPKRDPIQIEVYDIDLDGVDSPESCTGNQIARAL
ECCCHKNGESVITIGVAGQGEISTRPOLYTGWRWSAFEGVGRSELPSYVEMSER
GEIPLDTFITHMGLIEDINKAFDLMEGKSLRSYIH"
4804..5658
/locus_tag="pp1617"
4804..5658
/locus_tag="pp1617"
/note="identified by match to PFAM protein family HMM
PF00756"
/codon_start=1
/transl_table=1
/product="esterase, putative"
/protein_id="AA067238.1"
/db_xref="GI:24983099"
/translation="MSLDNISQCKSFGGWMKRYRHHSKVLGCDNVFVLYLPPQAGEGE
KLPLVWLSGITCTDENFMQAGQALAAELGLITVAPDSPREQVPGPDGMDG
LGAGFYLNATQOPWAQHYRMHDVVEELPALIEAHFPASGRSISGSHSGHALVCA

gene
/protein_id="AA067239.1"
/db_xref="GI:24983100"
/translation="MRIGHCYDVRFPDGPRTITGVRIKPKYGLLAHSDQVLLHAL
SDALLGALADSDIGKHPDDDPQKGRASRLNHYGYKAKKWKGNVDATTVAQA
PKMAPIHETMRQLLAEIDLQVELDQVNVKATTEKLGFTGREGIIVASHVALLPA"
6241..7299
/locus_tag="pp1619"
6241..7299
/locus_tag="pp1619"
/note="identified by match to TIGR protein family HMM
TIGR00094"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA067240.1"
/db_xref="GI:24983102"
/translation="MTELELLGPRASEPGLTALIKVAEDFOVDEVILIPLSGQGEH
LMLWVERDLNTEBEARLARAAAGVPYRISYAGIKROALTROMFSIHLPGKADPL
SRADASLRVLYKQVHORAKIQRGAHNSANGFTLRLLTALAADHQAVDARLEOLRQGVN
YFGTORHGGGVNDALDMARKALPEORVRVRSLSAGRSYVFNOLIAARVADGSW
ARAOVDILATDSRSRFPAGEACAPRIALIIDLHPGPMWGASBPAGAPOLEN
AIGARQPALCQMLAQAGMDHERILRLPTGGLTWHPPEPDLILEFVLPACCFATVY
REVYDLVASGQTDSPCVF"
7287..8036
/gene="sure"
/locus_tag="pp1620"
7287..8036
/locus_tag="pp1620"
/gene="sure"
/locus_tag="pp1620"
/note="similar to GP_8141682; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="stationary-phase survival protein Sure"
/protein_id="AA067241.1"
/db_xref="GI:24983103"
/translation="MRLISNDGCVTAAGIALHAALADYAECAVIAAPDQKSGASS
LTLRPLHPOTLANGFSLNCTPTDVCYHLGLNGLLPQMPDQVVGINLGARLGDVLY
SGTYAALIEGFLGCTSLARSLRSROPNLPTAATYARLVEADSRVLYPPRYLVAN
IPNLPLEHIRIGIQLTRUGHRARAAPTKVAVPVRKKEGVIWAVAGDAEDGGGTDFHAY
MGGVSTPPLDRTFNDAPFDLQDGMLEGLL"
8015..8710
/gene="pcm"
/locus_tag="pp1621"
8015..8710
/gene="pcm"
/locus_tag="pp1621"

Alignment Scores:
Pred. No.: 3,41e-05 Length: 305520
Score: 227.00 Matches: 133
Percent Similarity: 32.39% Conservative: 95
Best Local Similarity: 18.89% Mismatches: 264
Query Match: 6.91% Indels: 212
DB: 1 Gaps: 27

US-09-856-327-2 (1-618) x AE016780 (1-305520)

OY 29 GIuThrTyrgIyAsnaSpGlyValAspValPheIlealagIySerGlyProIleGlyAla 48
Db 47992 CAACCCAGAACATCGAAACCGCGATGTGTCGTGGTGGCCGCCGATGCGCGCAGT 48051

OY		49	ThrTyrAlauysLeucysValGluAlaGlyLeuArgValValMetValGluIleGlyAla	68
Dd		48052	GTCATGGCTACCAAGCATTGGGTGGCGGGGTGAATAACTGTGGTGCTGCATGGAGGCC---	48108
OY		69	AlAspSerPheTyrlAlaValAlaSnAlaGlUGlUGlYThrAlaValProTyrlAlaProGly	88
Dd		48109	-----CCGGCATTTCCA-----	48122
OY		89	TyrHisLysLysAengUllEglUpHeGLnLysAspIleAspArgPheValAlaSnValIle	108
Dd		48121	-----CCGAATCCACCAGCAGATAC-----CTGGAGCGCTTCTACACCGCGGTG	48162
OY		109	LysGlyAlaLeuGlnGlnValSerValProValArgAsnGlnSnValProThr-----	126
Dd		48163	CTCAAGGCGGCCGAGTCCGCGCTACCCACTGACAGAGGCGAGAAGACCCGGCCAGCAG	48222
OY		127	-----LeuAspProGlyAlaTrpSerAlaPropro	136
Dd		48223	TTTCGGCCACGCGGCGACCATTCACAGAACCTACACCTCACAGTAGTCCAAAGACCCCAACAAG	48288
OY		137	GlySerSerAlalIeserAngLyLysAnProHISGlnArgGluPheGlnsnLeuSer	156
Dd		48283	GCACGAAAAGCTCACTGTCGACAAAGAGGCCACAC-----CTGTTC	48322
OY		157	AlaGluAlaValThrArgGlyValGlyIymeserThrHisTrpThrCysserThrPro	176
Dd		48325	GCCAGCACTTAAGAGCGGTAGAGGGGGGACCATATGTCATGGATGGGACGGCCCTG	48384
OY		177	ArgIleHisProprometGluSerLeuProGlyIleGlyArgProLysLeuSerAsnAsp	196
Dd		48385	CGCCTG-----TTGCCACAGACTTCGCAATGGCGCACTGGCTGCATCAAC	48422
OY		197	ProAlaGluAspArgLysGluTrp-----AsnGluLeuTyrrSer	209
Dd		48427	GTCGGGGTGCAC-----TGCCCATCATGTTACAGCAGCATCTGCAGACCCGCTACTCGC	48477
OY		210	GluLlaGluArgLeuIleGlyThrSerThrLys-----	220
Dd		48478	CGGCCCGAGAGAGATGGGTGTCTCCGCGATGTGCTCCGACACTTACTGGGCGTG	48537
OY		221	GluPheAspGluSerIleArgHisThrLeu-----ValLeuArgSerLeuGlnAsp	237
Dd		48538	ACSTTCCCGCAGGGCTCAAAATACAGATGCATGGCATCCCGCGGTGGCTGGTCAGACAG	48597
OY		238	-----AlaTylLysAspArgGlnArg-----IlePhe	246
Dd		48598	GGCCTGGGCAAGCAGCTCACGGGCAACCGTGTACCAAGCCCGGAGAAGGGCCTGCAGCG	48657
OY		247	ArgProLeuProLeuAlaCysHisArgLeuLysAlaSnlarProGluTyrlValGluTrpHis	266
Dd		48658	AAGCGCAGTACCAGGTCTCACCTCAATGTCCGGCAAAACCCC---	48699
OY		267	SerAlaGluAsnLeuPheHisSerIleTyrlAsnAspArgLysGlnLysLysLeuPheThr	286
Dd		48700	-----GCGGGGGCCCACTCGCAAAATGATRACGGCCGTGG-----GTG	48738
OY		287	LeuLeuThrAsnHisArgCystrHArgLeuAla-----LeuThrGlyLysTyrglu-----	303
Dd		48739	TGCCTGGCAACACCACTGCACGCGCATGTGCGCGCATTCACAGGCCAAAGTACAGACGCCAC	48799
OY		304	-----LysLysIleGlyAlaAlaGluValArgAsnLeuLeuAlaThr-----	317
Dd		48799	GTGACCATGGCCAAAGCACTGTGATACCGCAACGTGCGATCTCTTACACAGACCGTTGCA	48858
OY		318	-----ArgAsn	319
Dd		48859	AGCCAACTGCAGTGCATGCCAAATGGGGTACGCGGATCGATTTCATCAATCACAGAAC	48918
OY		320	ProSerSerGlnLeuAspSerTyrlIleMetAlaLysValTyrlValLeuAlaserylAla	339
Dd		48919	AATGTATCGGCCCAAGACGGCAGCGGCCCAAGCGCATCTGATTCGCCGCCCAAGCGG	48978
OY		340	IleGlyAsnProGlnIleLeuTyrlAsnSerGlyPheSerGlyLeuGlnValThrProArg	359

Db	48979	ATCGAATACACCGAAGCTGCTGCTCACTCC-----ATCGGCCCGGAAC	49020
QY	360	AsnaSperLeuIIleProAsnLeuGlyArgTyrIleThrfGluInProMetaIaphcys	379
Db	49021	AGCCCTAAGCGGTGGCCACACGACGAGCGCCAGGTGCGCCAG-----	49065
QY	380	GlnIleValLeuArgGlnGluPheValAspSerValArgAspAspProTyrGlyLeuPro	399
Db	49063	-----GCATTGCTGACCCCGCGGTGACTCGGCC	49092
QY	400	TrpTrpIlySgluAlaValAlaGlnHisIleAlaLysAspProThrAspIaleuProIle	419
Db	49093	TGG-----GCCCTGATGCGCCGAAGCAAGCCGCTGTTCC	49125
QY	420	ProPheArgAspPro-----GluProGlnValThrThrProPhe---	432
Db	49126	CGGTTCGCGGGCGCGTGTGTACTCGCGGAGATGAGACATGGCGCATGGCCGTTCCGC	49185
QY	433	ThrGlnGluHisProThrPrhIsthrGlnIleHisArgAspIaPheSerTyrGlyAlaVal	452
Db	49186	AGCCAGCGCTGCGCGCTGCGATGCAATCGCACAGACGGCGCTGGAACCTGGTGAAAC	49245
QY	453	GlyProGluValAspSerArgValIleValAspLeuArgTrpPheGlyAlaThrAspPro	472
Db	49246	GACCCCTACACACTTCGCTGTGATTTCATGACGGCACCAACACAGCGCGCGCTAACCCC	49305
QY	473	-----GluAlaAsnAsnLeuValPhe	480
Db	49306	AGCAATCTGGCGTGTGTCGGCGACAGCGCGCTGTGCAGACAGTCAACAGCGTGGACCGCC	49365
QY	481	Gln-----	481
Db	49366	CAGTTCCGCTGGCTTCTTGTCGACAGGTGAGAAAGCCCGATATGCGCTGTGC	49425
QY	482	-----AsnaSperValGlnAspGlyTyrSerMetProGlnProThrPheArg	496
Db	49426	CGGTGAGAGAGAGATGACGAGATTCTCCGACGAGATTGGGCTGCTCGCGCCGACGATCAC	49485
QY	497	TyrArgProSer-----ThrIaSerAsnValArgIaArgIysMetMetaIaAspMet	514
Db	49486	TACGAACTGTCCGACTACACCAAGAGGGTTTCAGACAGCGCGCGCTGACTACATAC	49545
QY	515	CysGluValAlaSerAsnLeuGlyGlyTyrLeuProThrSerProGlnPheMetAsp	534
Db	49546	ATC---ATCACCGAGCTGATGGCGGCTACCCGACCTGCAACCTTAAGCTCGAG	49602
QY	535	ProGlyLeu-----AlaLeu	539
Db	49603	CCCCGGCTCGCAACCGGTGTTGAGTACGACGCGCCAGACTACCGCTTACGGCGCTGCG	49662
QY	540	HisLeuAlaGlyThrThrArgIleGlyPheAspLysAlaThrThrValAlaAspAsnAsn	559
Db	49663	CACCTGAATGGCGCACTACCGCATGGGCTCGGACCGCACCAAGTGGGTGGACAAAGAC	49722
QY	560	SerLeuValTrpAspPheAlaAsnLeuTyrValAlaGlyAsnGlyThrIleArgTyrGly	579
Db	49723	CAGGCAAGCTGGGAGCACCCCAACCTGTCTGTGTCGGGACGCGCTGTTCCCACTACC	49782
QY	580	PheGlyGluAsnProThrIleuThrIleuThrSerMetCysHisAlaIleLysSerAlaArgSerIle	599
Db	49783	GGCACCGCCAAACCGGACCTTGACCATACACCGCGTGAAGTTCAGGCGCCGACAGCTGTG	49842
QY	600	IleAsnThrIleu	603
Db	49843	GCCAGCGCACTG	49854
RESULT 15			
LOCUS	AF430844	2465 bp	DNA linear BCF 26-MAR-2003
DEFINITION	Burkholderia cepacia glucose dehydrogenase (gdhAlpna) gene, complete cds.		
ACCESSION	AF430844		


```
Db      1791 TCCTGATCGTTTCGGACGCTCCGTCGCGCGACCGAAGCGCGAAGATCCAC 1850
QY      444 -----ArgAspAlaPheSerTyrGlyAlaVal 452
Db      1851 CTGTGCAACCTGTCCGCGATCGACCGAGACGACGACAGATCTTCAAGCCGCAAGCTG 1910
QY      453 -----GlyProGluValAspSerArgVal-----IleValAspLeu 464
Db      1911 ATGAAGCCCGACGACGCTCGACCGAGATCCGCGACCGCTACGTCGACGATTG 1970
QY      465 ArgTTPPheGlyAlaThrAspProGluAlaAsnAsnLeuValPheGlnAsnAspVal 484
Db      1971 GACTGCTTCACGAAATCCGCGCAACCGCAACCGCATGTCGCCGACAAGACGGCG 2030
QY      485 GlnAspGlyTyrSerMetProGlnProThrPheArgTyrArgProSerThrAlaSerAsn 504
Db      2031 ACCGATGGCATTCGCGATTCGCGCCCGCGATCAGCTAT-----GCGATCGAC 2078
QY      505 ValArgAlaArgLysMetMetAlaAspMetCysGluVal-----AlaSerAsn 520
Db      2079 GACTACGTGAAGCGCGCGCGCGCATACGCGAGGTCTACGCGACCGCGCGAAGTGT 2138
QY      521 LeuGlyGly-----TyrLeuProThrSerProProGlnPhe 532
Db      2139 CTCGGCGCGACGACGCTCGTTCAACGACGAATTGCGCGCGACACAAAT----- 2186
QY      533 MetAspProGlyLeuAlaLeuHisLeuAlaGlyThrThrArgIleGlyPheAspLysAla 552
Db      2187 -----CNCATCACGGGCTCGACGATCATGGCGCGCATGCGCGC 2225
QY      553 ThrThrValAlaAspAsnAsnSerLeuValTyrAspPheAlaAsnLeuTyrValAlaGly 572
Db      2226 GACTCGTCGTGACACAGGACTGCCGACAGTTCGACCATCGAACCCTGTCATTTCGAGC 2285
QY      573 AsnGlyThrIleArgThrGlyPheGlyGluAsnProThrLeuThrSerMetCysHisAla 592
Db      2286 ACCGCGACGATGCCGACGCTGTAACCGTAACGTAACGTAACGTAACGTAACGTAACGTA 2345
QY      593 IleLysSerAlaArgSerIleIleAsnThrLeuLys 604
Db      2346 CTGCGGATGTCG-----GACACGCTGAAG 2369
```

Search completed: September 24, 2003, 07:57:18
Job time : 6746 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 01:45:38 ; Search time 7651 Seconds
(without alignments)
11260.708 Million cell updates/sec

Title: US-09-856-327-1
Perfect score: 2106
Sequence: 1 atcagccatgctctctca.....aaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Geneml:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_hgo_hum:*

40: em_hgo_mus:*

41: em_hgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2106	100.0	2106	6 BD013176	BD013176 A novel p
2	225.8	10.7	2069	8 AY291124	AY291124 Trimetes
3	221	10.5	2064	8 AF353193	AF353193 Peniphor
4	212.4	10.1	1869	6 AR141573	AR141573 Sequence
5	210.2	10.0	1869	6 E117166	E117166 cDNA of pyr
6	210.2	10.0	1869	6 I82410	I82410 Sequence 1
7	210.2	10.0	2046	8 D73369	D73369 Coriolus ve
8	175.6	8.3	1902	6 E32737	E32737 Panus cornu
9	152.6	7.2	1701	6 AR170485	AR170485 Sequence
10	129	6.1	1946	6 E37379	E37379 Novel prote
11	129	6.1	1946	6 E37379	BC032271 Reagent for
12	129	6.1	1946	8 AB043883	AB043883 Tricholom
13	94.6	4.5	2013	9 AB070201	AB070201 Macaca fa
14	94.4	4.5	589	10 BC049706	BC049706 Mus muscu
15	94.4	4.5	1901	5 BC048876	BC048876 Danilo rer
16	94.4	4.5	4640	6 AX376008	AX376008 Sequence
17	93.8	4.4	3005	10 BC032271	BC032271 Mus muscu
18	92	4.4	2360	10 BC008547	BC008547 Mus muscu
19	92	4.4	2578	9 AB049758	AB049758 Homo sapi
20	92	4.4	3660	10 BC051069	BC051069 Mus muscu
21	91.6	4.3	1618	3 AY069072	AY069072 Drosophila
22	91.6	4.3	1624	3 AK114250	AK114250 Clona int
23	91.4	4.3	1911	10 BC048954	BC048954 Mus muscu
24	91.2	4.3	2091	6 BD176856	BD176856 A method
25	91.2	4.3	2091	17 AF111849	AF111849 Homo sapi
26	91.2	4.3	2964	6 AX574582	AX574582 Sequence
27	91	4.3	1282	9 BC017745	BC017745 Homo sapi
28	90.8	4.3	845	9 BC044653	BC044653 Homo sapi
29	90.8	4.3	1327	6 AX306545	AX306545 Sequence
30	90.8	4.3	1817	10 BC051174	BC051174 Mus muscu
31	90.6	4.3	1567	9 AB047887	AB047887 Macaca fa
32	90.6	4.3	2537	9 BC003684	BC003684 Homo sapi
33	90.4	4.3	2335	9 AB055368	AB055368 Homo sapi
34	90.4	4.3	3983	9 HSM603572	HSM603572 Macaca fa
35	90	4.3	1289	9 BC032350	BC032350 Homo sapi
36	90	4.3	2384	3 AY069709	AY069709 Drosophila
37	90	4.3	2566	17 AF130082	AF130082 Homo sapi
38	90	4.3	3809	9 HSM604486	HSM604486 Homo sapi
39	89.8	4.3	470	3 AF177908	AF177908 Hydra vul
40	89.8	4.3	1864	9 AB056421	AB056421 Macaca fa
41	89.8	4.3	2180	8 AB028665	AB028665 Gentiana
42	89.8	4.3	4861	9 BC040342	BC040342 Homo sapi
43	89.6	4.3	1286	9 BC044234	BC044234 Homo sapi
44	89.6	4.3	2032	9 AK027204	AK027204 Homo sapi
45	89.6	4.3	2260	6 BD021751	BD021751 Novel gen

ALIGNMENTS

RESULT 1
BD013176
LOCUS
DEFINITION A novel protein, a gene coding therefor and a method of using the same.
ACCESSION BD013176
VERSION BD013176.1 GI:22093365
KEYWORDS
SOURCE WO 0121657-A/1.
ORGANISM Lyophyllum shimeji
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Tricholomataceae; Lyophyllum.
REFERENCE 1 (bases 1 to 2106)
TAKAKURA,Y., Kuwata,S. and Inoue,Y.
TITLE A novel protein, a gene coding therefor and a method of using the

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0121657-A 1 29-MAR-2001;
JAPAN TOBACCO INC., CORPORATE JURIDICAL PERSON SOCIETY FOR TECHNICAL
INNOVATION OF AGRICULTURE YUKIO SHIMOMURA, SHIRO TAKEKAWA, NOBO CHO
FORESTRY AND FISHERIES, YOSHIMITSU TAKAKURA, SHIGERU KUMATA, YASUHIRO
INOUE
OS Lyophyllum shimeji
PN WO 0121657-A/1
PD 29-MAR-2001
PF 20-SEP-2000 WO 2000JP006404
PR 21-SEP-1999 JP 99P 267238
PI YOSHIMITSU TAKAKURA, SHIGERU KUMATA, YASUHIRO INOUE PC
C07K14/375, C12N15/31, C12N15/63, C12N1/21, C12Q1/68, C12P21/02, PC
A01M65/00
CC
FH Key Location/Qualifiers
FT CDS (8)..(1861).
FEATURES
source 1..2106 location/Qualifiers
/organism="Lyophyllum shimeji"
/mol_type="genomic DNA"
/db_xref="taxon:47721"
BASE COUNT 593 a 540 c 527 g 446 t
ORIGIN
Query Match 100.0%; Score 2106; DB 6; Length 2106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCAGCCATGCTCTCTCAACGAGCAGATGCTAGCGGCTATCCAGCGGTATGCAAT 60
DB 1 ATCAGCCATGCTCTCTCAACGAGCAGATGCTAGCGGCTATCCAGCGGTATGCAAT 60
QY 61 CAACGACAGATTTCTTAAGAACGCAATTCAGAAACATACGGAACGAGAGTGTAT 120
DB 61 CAACGACAGATTTCTTAAGAACGCAATTCAGAAACATACGGAACGAGAGTGTAT 120
QY 121 ATTCAATGAGATCTGACACCATTTGAGCGACGATATGCAAAAGCTCTGTGTAAGCTGG 180
DB 121 ATTCAATGAGATCTGACACCATTTGAGCGACGATATGCAAAAGCTCTGTGTAAGCTGG 180
QY 181 TCTAGTGTGTGATGATGATGAGATGAGACCTGCTATAGCTTCTACGCTTAAATGCCGA 240
DB 181 TCTAGTGTGTGATGATGATGAGATGAGACCTGCTATAGCTTCTACGCTTAAATGCCGA 240
QY 181 TCTAGTGTGTGATGATGATGAGATGAGACCTGCTATAGCTTCTACGCTTAAATGCCGA 240
DB 181 TCTAGTGTGTGATGATGATGAGATGAGACCTGCTATAGCTTCTACGCTTAAATGCCGA 240
QY 241 AGAAGAACTGCAAGTTCCTACGTTCTGCGTACCAACAAGAAATGAATGAGATTCCA 300
DB 241 AGAAGAACTGCAAGTTCCTACGTTCTGCGTACCAACAAGAAATGAATGAGATTCCA 300
QY 301 GAAAGATATGACCGCTTCTGCAATGTATCAAGGAGCCTTACAAAGTCTCTGTTCC 360
DB 301 GAAAGATATGACCGCTTCTGCAATGTATCAAGGAGCCTTACAAAGTCTCTGTTCC 360
QY 361 TGTCAAGAAACGAGACGTGCTACACTTGATCCGAGAGCTGAGCGGCCCCCTGGAG 420
DB 361 TGTCAAGAAACGAGACGTGCTACACTTGATCCGAGAGCTGAGCGGCCCCCTGGAG 420
QY 421 TTCAGCCATATGAAAGGTAATAATCTCACCAAGCGGGAATTCGAAATTTGTCGGA 480
DB 421 TTCAGCCATATGAAAGGTAATAATCTCACCAAGCGGGAATTCGAAATTTGTCGGA 480
QY 481 GGGCCATAACGCGTGGAGTCGGGCGCATAGTACCCACTGTGAGCTCTCCACCCAGGAT 540
DB 481 GGGCCATAACGCGTGGAGTCGGGCGCATAGTACCCACTGTGAGCTCTCCACCCAGGAT 540
QY 541 TCATCCAGCCATGAAAGTCTCCGGGCGATCGGCGCTCGAAAGCTCAGTAACGACCCGGC 600
DB 541 TCATCCAGCCATGAAAGTCTCCGGGCGATCGGCGCTCGAAAGCTCAGTAACGACCCGGC 600
QY 601 AGAGAGCAACAAGAGTGAAGCAAGCTTTATTCGAGGCGGAGCTCTCATCGGAGCTTC 660
DB 601 AGAGAGCAACAAGAGTGAAGCAAGCTTTATTCGAGGCGGAGCTCTCATCGGAGCTTC 660
QY 661 CACCAAGAAATTCGAGAGTCAATTCGGGACACCCCTTGTCTGCGCTCTTTCGAAGAGC 720

DB 661 CACCAAGAAATTCGAGAGTCAATTCGGGACACCCCTTGTCTGCGCTCTTTCGAAGAGC 720
QY 721 GTACAAGAGTGTCAACGATCTTTGCGCCCTCCCGTGGCATGCGACCGGTTGAAGA 780
DB 721 GTACAAGAGTGTCAACGATCTTTGCGCCCTCCCGTGGCATGCGACCGGTTGAAGA 780
QY 781 CGCGCGGAATACGTGGAATGCGACTCAGCAGAAAAATCTTTTCACTATCTATACAGA 840
DB 781 CGCGCGGAATACGTGGAATGCGACTCAGCAGAAAAATCTTTTCACTATCTATACAGA 840
QY 841 TCACAAGCAAGAAGCTTTTACCGTGTGACGAACCATTCCTGTGACACGACTGGCGT 900
DB 841 TCACAAGCAAGAAGCTTTTACCGTGTGACGAACCATTCCTGTGACACGACTGGCGT 900
QY 901 TACGGGCGGATGTGAAGAAGATTTGCGCCCTCCGAGGTCAGGAATCTCTGCGCACGC 960
DB 901 TACGGGCGGATGTGAAGAAGATTTGCGCCCTCCGAGGTCAGGAATCTCTGCGCACGC 960
QY 961 GAATCCTATGTCGACGTGAGCAGCTATATCATGAGGGAAGATATGATCTGCGCGG 1020
DB 961 GAATCCTATGTCGACGTGAGCAGCTATATCATGAGGGAAGATATGATCTGCGCGG 1020
QY 1021 ACGGATCGGCAACCCACAGATTTCTATATCTGCGGCTTCTGTGGCTATCAGCTACGCC 1080
DB 1021 ACGGATCGGCAACCCACAGATTTCTATATCTGCGGCTTCTGTGGCTATCAGCTACGCC 1080
QY 1081 ACGCAATGACTGTTGATCCCACTGGGAGGTATACATCAGGAGACAGCCGATGCAAT 1140
DB 1081 ACGCAATGACTGTTGATCCCACTGGGAGGTATACATCAGGAGACAGCCGATGCAAT 1140
QY 1141 TTGCAGATAGTCTTGAGCAGGAATTCGTCAGACGCTGCGCAGATCTTATGACT 1200
DB 1141 TTGCAGATAGTCTTGAGCAGGAATTCGTCAGACGCTGCGCAGATCTTATGACT 1200
QY 1201 GGCATGTTGGAAGAAGCCGTTGCTCAACATTTGCGCAAGAACCAGAGATGCACTGCC 1260
DB 1201 GGCATGTTGGAAGAAGCCGTTGCTCAACATTTGCGCAAGAACCAGAGATGCACTGCC 1260
QY 1261 CATTCGTTCCGCGATCCGGAACCCAGGTATCAACCCCATTTACAGAAAGACCCCTG 1320
DB 1261 CATTCGTTCCGCGATCCGGAACCCAGGTATCAACCCCATTTACAGAAAGACCCCTG 1320
QY 1321 GCACAGCAGATTCACCGGATGCTTTTCTGAGGTCGCTGCTGAGTGGACTC 1380
DB 1321 GCACAGCAGATTCACCGGATGCTTTTCTGAGGTCGCTGCTGAGTGGACTC 1380
QY 1381 TCGTGTCAATCGTCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 TCGTGTCAATCGTCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1441 GGTTCGAGAAAGATGTTTCAAGAGCGGATAGTATGCGCGAGCGGATGATGATG 1500
DB 1441 GGTTCGAGAAAGATGTTTCAAGAGCGGATAGTATGCGCGAGCGGATGATGATG 1500
QY 1501 ACCGAGCACTGCTCAAAAGTGAAGCAAGGAAATGATGCGGATATGTCGAGTGGC 1560
DB 1501 ACCGAGCACTGCTCAAAAGTGAAGCAAGGAAATGATGCGGATATGTCGAGTGGC 1560
QY 1561 GAGCAACTGGGAGGTTATTTGCCACGTCGCCCGGAGTTATGATGATCAAGGCTTGC 1620
DB 1561 GAGCAACTGGGAGGTTATTTGCCACGTCGCCCGGAGTTATGATGATCAAGGCTTGC 1620
QY 1621 ACTTCATCTTGGGGGACTACTCGCATTTGCTTCAGCAAGCAATACAGTGGCTGATTA 1680
DB 1621 ACTTCATCTTGGGGGACTACTCGCATTTGCTTCAGCAAGCAATACAGTGGCTGATTA 1680
QY 1681 CAATCGCTGTGTGGGACTTTTGCATCTTTATGTCAGGCAATGCGACCATCAGAC 1740
DB 1681 CAATCGCTGTGTGGGACTTTTGCATCTTTATGTCAGGCAATGCGACCATCAGAC 1740
QY 1741 GGGCTTGGGAGAACCCGACACTTACGTGATGTCACAGCTATCAAGAGGCGCAGAG 1800

Db 1741 GGGCTTGGCGAGAACCCGACCTTACGTGATGTGCCAGCTATCAAGAGCGGAGAG 1800
QY 1801 CATCATCAATACACTCAAGGTGGGACTGACGAGAAAAATATCAGGCGAGCTCCCAACT 1860
Db 1801 CATCATCAATACACTCAAGGTGGGACTGACGAGAAAAATATCAGGCGAGCTCCCAACT 1860
QY 1861 TTGAGAGAGGAGCAACAGCAGTGTAAACAAACGGCTCAAGTGGCTACTTCAAGTTGAATG 1920
Db 1861 TTGAGAGAGGAGCAACAGCAGTGTAAACAAACGGCTCAAGTGGCTACTTCAAGTTGAATG 1920
QY 1921 CATCTGGTCCCTACCATATGTATGTATGTATGACATAGCGCTTGAAGATTGTATATAC 1980
Db 1921 CATCTGGTCCCTACCATATGTATGTATGTATGACATAGCGCTTGAAGATTGTATATAC 1980
QY 1981 TGAACCTGTACTTGTCTGATATAGTATGACATATGATCATGTTTAAAAAATTTTAC 2040
Db 1981 TGAACCTGTACTTGTCTGATATAGTATGACATATGATCATGTTTAAAAAATTTTAC 2040
QY 2041 TGAACCTGTACTTGTCTGATATAGTATGACATATGATCATGTTTAAAAAATTTTAC 2100
Db 2041 TGAACCTGTACTTGTCTGATATAGTATGACATATGATCATGTTTAAAAAATTTTAC 2100
QY 2101 AAAAAA 2106
Db 2101 AAAAAA 2106

RESULT 2
AY291124 2069 bp mRNA linear PLN 23-MAY-2003
LOCUS AY291124
DEFINITION Trameetes ochracea strain MB49 pyranose oxidase (p20) mRNA, complete cds.
ACCESSION AY291124 GI:31044223
VERSION AY291124.1 GI:31044223
KEYWORDS
SOURCE Trameetes ochracea
ORGANISM Trameetes ochracea
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Aphyllophorales; Trameetes.
REFERENCE 1 (bases 1 to 2069)
Vecerek, B., Maresova, H., Kocanova, M. and Kyslik, P.
Molecular cloning and expression of the pyranose-2 oxidase gene from Trameetes ochracea MB49 in Escherichia coli
Unpublished
JOURNAL 2 (bases 1 to 2069)
Vecerek, B., Maresova, H. and Kyslik, P.
Direct Submission
AUTHORS Submitted (06-MAY-2003) Institute of Microbiology AS CR, Videnska
TITLE 1083, Prague 4 142 20, Czech Republic
JOURNAL location/Qualifiers
FEATURES
source
1..2069
/organism="Trameetes ochracea"
/mol_type="mRNA"
/strain="MB49"
/db_xref="taxon:230624"
1..2069
/gene="p20"
49..1920
/gene="p20"
/codon_start=1
/product="pyranose oxidase"
/protein_id="AAP40332.1"
/db_xref="GI:31044224"
/translation="MSSSSSDPPEFPAKSSFRSAANOKASASSLPRLGDKKVPKND
IKYDVIYVSGPTIGCTYARBLVAGRYVAAGVADGELIDSGIKGHNKNTYBYKNDIK
EVNIDGQLSVSPVNTLVVDLSPSWASTFEVNGSNPEDDPLRLNLSGAVTK
VGSMTHMTCATPRDFEORPLVTKDADADAEMLVTKABSYFQGTGDOFESIR
HNVLNKLTEYRGORDFQIPLAATRRSPTEVMSANTVFDLQNPNDAPERN
LPPAVACERVVRLNLSLHDLISDREIKADYVVLGAGAHNQLVNSGE
GOLGRPNAPPELPLSISYITEOSLYRCQVSTPLDSVSDMTIRGTBGLTYS
VYTPPASTKHKHDMNEKYNHMQDEPLPLPPEDEPOVYTTLEQPSHWHITQIH
RDAPSTGAVDQSDISRLIVMREFGRTPEENKLMVSDKITDAVYNNPOTPFERFPA
GRSTKEADMDMTDMCVMSAKIGGFLPGLPGLVLIHGLGTHRMGPFDEKEDNCV

BASE COUNT 449 a 659 c 584 g 377 t
ORIGIN
Query Match 10.7%; Score 225.8; DB 8; Length 2069;
Best Local Similarity 51.6%; Pred. No. 4.3e-36;
Matches 842; Conservative 0; Mismatches 662; Indels 129; Gaps 9;
QY 268 TGGCTACCAAGAAAGATGAATGAGTTCAGAAAGATATGACGGCTTCGTCATATG 327
Db 306 TGGTCCACAGAGAGAACACCGTGCAGTACCAAGAACATGCACAAATTTTGAACGT 365
QY 328 AATCAAGGAGACCTTACAAACAGTCTGTCTGTCTGTGAAGAACGAGACCTTACACT 387
Db 366 CATTCAGGAGAACATGATGATGTGTTCGTTCCGCTCAATATCCCTGTGGTGCAGACGT 425
QY 388 TGAATCCCGAGCCTTGAGACCGGCCCTCGAAGTTACACCATATGCAACGCTAAAAATCC 447
Db 426 CAGCCGACGCTCTTGGCAAGCTTCGACG-----TTCTTCGTCGCAACGCTCGAACCC 479
QY 448 TCACGAGCGGAATTCGAACCTTGTCGCGGAGCGCGTAAAGCGGTGAGTCGCGGACAT 507
Db 480 AGACGAGGAGACCCACTTGTGTAACCTTCAGTGTGACGCGCTCAGCGCTGTCGAGACAT 539
QY 508 GAGTACCCACTGAGACGTCTCCACGCGCACGCGATTCATCCACCATGGAAGTCTCCCGGG 567
Db 540 GTTACGACACTGAGACATGCGCCACACCCCGTTTGA-----CCGGGA 581
QY 568 CATGCGCCGTCGGAAGCTCACTAACGACCGCGCAGAGAGACGACAAAGTGGAAACGACT 627
Db 582 GCMAAGCCCGCTGCTTGTGAAGAGACGCGACGCTGACACCTGTGTCGAGGACCGGCT 641
QY 628 TTATTCGAGGCGGAGGCTCATCGGGACTTCCACCAAGAAATTCGAGTGAATTCG 687
Db 642 CTACACCAAGGCGCGAGTGTACTTTCACAGTGTGACGCGACCATTTCAAGATGATTCG 701
QY 688 GCACACCCCTTGTCTGCGCTCTTTCGAAGACGCGCTACAGAGATGCTAACCTATCTTCG 747
Db 702 CCACAAATCTGTCTGTCACAAAGCTCAGAGAGAAATPACAGGCGCAAGCGCATTCAGCA 761
QY 748 CCTCTCCCGTTGGCATGCGACCGGTTGAAGAACGCGCGGAAATACGTGATGAGCACTC 807
Db 762 GATTCACACTGCGCGCAACCGCGCG-----AGCCCGACCTTCGTAATGAGACTC 812
QY 808 AGCAGAAATCTTTTCCACTTATCTAACAGATGACAGAGAGAAAGCTTTACCT 867
Db 813 GCGCAACACTGTTTT-----CGACTTCAGAACAGCGCGAACACGAGCGC 857
QY 868 GCTGACGACCATGCTGCACACGACTGCGCTTACGGGCGGTATGAGAAGAAATTCG 927
Db 858 GCGGAGAGAGGCTTCAACCTCTTCCCGCGGTGCGGTGTAGCGCGTCTGTGCGCAACGC 917
QY 928 CGCTGCCGAGTCAAGAACTACTGCGCACCAAGGAATCTTATGTCGACGTGACACACTA 987
Db 918 GTTGAACCTGAGATGACAGATCTGCACATTCACGACCTCATCTCGGGCGACCTTCGA 977
QY 988 TATCATGCGAAGATATATGATGCTGGGTGGGAGGATGCGCAACCAACCAATTCCTA 1047
Db 978 AATCAAGCTGACGTGTACGCTCTACCGCGCGGGGCTTCACACACGCGCTTCTGCT 1037
QY 1048 TAATCGGGCTT-----CTCTGGGCTACAGGTGTCGCCACGCAAGTACTGTTGATCCC 1101
Db 1038 GAATCTGGCTTTGACAGAGCTGGGCGCGACGCAACCCGCAACCCAGCGAGCTGTGCGC 1097
QY 1102 CAACCTGGGAGATCATACAGGACAGCGAGATGCAATTTTGCCAGATPAGTCTTGAGGCA 1161
Db 1098 GTCCCTGGGAGGCTACATCACCGAGAGTGTCTTGTGCGACAGCGGATGAGACAC 1157
QY 1162 GGAATTCGTGACAGCGT----- 1179
Db 1158 CGAGCTATGACAGCGCTCAAGTCCGACATGACCATATGAGGGAACCCCTGGCGAGCTGAC 1217
QY 1180 -----GCGGACGATCTTATGGA-CGTGCAATGTTGGAAGA 1215

```

Db      1218 GTACAGCGTACGTACACGCCGCCGCTGCAGCAACAGCACCCGAGCTGGTGGAACA 1277
QY      1216 AGCCGTGCTCAATATTGTCACAGAACCCGACAGATGACATGCCATTCGGTCCGCGA 1275
Db      1278 GAAGGTGAAMACACATATGACAGACACAGAGAGCCCGCTCCCATCCGTTCCAGGA 1337
QY      1276 TCCGGAACCCAGAGTAAACACCCATTATAGAGAAGAACCCCTGGCACAGCAGATTGA 1335
Db      1338 CCCCGAGCCGAGAGTACCATCTGTTTCAACCGTGCACCCGTCACACATCAGATCA 1397
QY      1336 CCGGATGCTTTTGTACGGTGCCTGCTGAGTGCAGTCTCGTGTATCTGCA 1395
Db      1398 CCGCGATGCTTTCAGTTACGGGCACTGACGAACACATGACCTCGCTCATCTGTGA 1457
QY      1396 CCGCGATGCTTTCAGTTACGGGCACTGACGAACACATGACCTCGCTCATCTGTGA 1455
Db      1458 CTGGCCCTCTTTCAGTTACGGGCACTGACGAACACATGACCTCGCTCATCTGTGA 1517
QY      1456 TGTTCAGACGGGTACAGTATGCGCGACGCCGATTCAGATATCG--ACCGACACTGC 1512
Db      1518 GATCAGGACGGGTACAGTATGCGCGACGCCGATTCAGATATCG--ACCGACACTGC 1577
QY      1513 GTCAACGTTAGACAGAGAAATGATGGCCGATATGTCGAGTGCAGAGTGCAGCACTTGG 1572
Db      1578 CAGGACGAAAGAGGCGGAGACATGATGACGACATGCTGTTATGTCGCGAAGATCGG 1637
QY      1573 AGGTTATTTGCCACGTCCTCCCGCGATTTATGATCCAGGCTTCGACTTCATCTTGC 1632
Db      1638 TGGCTTCCTACCCGCTCCCTCCCGATTCATGAGGCTTGGTCTTGTCTTACCTTGG 1697
QY      1633 GGGGACTACTCCGATTTGGCTTGAC-----AAGCACTACAGTGGCTGATPAACAATC 1686
Db      1698 TGGTATGCAACGCGATGCTGACGAGAGAGCACTGCTGCTCAACAACGACTTC 1757
QY      1687 GGTGGCTGGGACTTTGGCATCTTATGTTGACGACATGGACACCATCAGACGGGCTT 1746
Db      1758 GCGCGTGTTCGCTTCAAGACCTCTCTCGTGCTGCGGAAATTCCTTACCGCGTA 1817
QY      1747 CCGGAGAACCCGACACTTACGTCGATGTCACGCTATCAGAGGCGGAGGACATCAT 1806
Db      1818 CCGGAGAACCCGACGCTACCGCATGTCGCTCGGATCAAGAGTGGGAGTACATCAA 1877
QY      1807 CAATACACTCAAG 1819
Db      1878 GCAGAACTTCAG 1890

RESULT 3
AF535193      2064 bp      mRNA      linear      PLN 31-DEC-2002
LOCUS      Peniophora sp. SG pyranose oxidase mRNA, complete cds.
DEFINITION      AF535193
ACCESSION      AF535193.1 GI:27436421
VERSION
KEYWORDS
SOURCE      peniophora sp. SG
ORGANISM      Peniophora sp. SG
REFERENCE      1 (bases 1 to 2064)
AUTHORS      Heckmann,D.M., Bastian,S., Rekowski,M. and Giffhorn,F.
TITLE      Pyranose Oxidase of the White-Rot Fungus Peniophora sp. Strain SG:
            Cloning and Characterization of the Gene, Heterologous Expression
            in Escherichia coli, and Properties of the Recombinant Enzyme
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 2064)
TITLE      Heckmann,D.M. and Giffhorn,F.
AUTHORS      Direct Submission
JOURNAL      Submitted (07-AUG-2002) Applied Microbiology, University of the
            Saarland, Im Stadtwald, 66123 Saarbruecken, Saarland, Germany
FEATURES
            source
            1..2064
            /organism="Peniophora sp. SG"

```

```

CDS
/mol_type="mRNA"
/strain="SG"
/db_xref="taxon:204723"
26..1897
/EC_number="1.1.3.10"
/note="FAD-oxidoreductase; pox; pyranose: oxygen
2-oxidoreductase; poxSG"
/codon_start=1
/product="pyranose oxidase"
/protein_id="AA013382.1"
/db_xref="GI:27436422"
/translation="MSTSSDPFPFAKSFPSAARASASLPIRLPDKKVGMD
IKDIVVSGIPIGCTVARELVAGIKVAMFIDIGLSLKGARKNTVEROKIDK
EVNVIQGLMSVSPVNTLVDTLSPTSWADSTFVRNGSNPEODPLRLSGOAVTR
VGMSTHMTCATPRFDRQRLPLVDDADDAEDMDRLYTKAESYFGCTDOFKESIR
HNLVINKLAEEKYGGORFOIPLAATRRSPFEVMSANVTVDLONRPTDPERFN
LEPAVACERVVKNALNSETESLHIDHLSGDRFELKADYVLTACAVHNTOLVNSG
GOLGRNPTNPPELLPLSGSYITTESLYFCQVYMETELIDSKSWITRGTPGLTYS
VYTFGASTNKHHPDMWNEKVKHNMQHODPLIPDEDEPVVTLFQSPHMHQI
RDAFSYGAVOQSIDSLIVDMRFEGTERTERENKLMFSDKITDAYNMQPTDFRPA
GRTSKAEEDMMDKCVMSAKIGFPLGSLPQPMELGLVHLGSTRMGFDEKEDNCV
NDSRVFGRKNIPLGCGNIPYAGVANTLITAMSLAIKSCVEYIKQNTFPTPSBAO"
poly_site      2047
BASE COUNT      462 a      652 c      579 g      371 t
ORIGIN
Query Match      10.5%; Score 221; DB 8; Length 2064;
Best local Similarity 51.4%; Pred. No. 4.2e-35;
Matches 839; Conservative 0; Mismatches 665; Indels 129; Gaps 9;
QY      268 TGGCTACACAGAAATGAAATGAGTTCACGAAAGATATTGACCGCTTGCAATGT 327
Db      283 TGGTGCCACAGAAAGAACCCGTCGATGACGAGAAACATGCAAGTTGTACACT 342
QY      328 AATCAAGAGGCTTACACAAAGTCTCTTCTGAGAAACGAAAGTGCCTACACT 387
Db      343 CATTCAGGACATTTGATGTCGTTCCGTCCTGATACCTCGTGGTGAACAGCT 402
QY      388 TGATCCCGAGCGCTGAGCGCCGCCCTTGAAATTCAGCATATGAAAGGTAAATTC 447
Db      403 CAGCCGACGCTGTGGCAAGCTTCGACG-----TTCCTGTCGCCCAACGCGCTCAACC 456
QY      448 TCACACGCGGGAATTCGAAACTTGTCTGCGAGGCCGTAAACGCTGGAGGCACT 507
Db      457 AGAGGAGGACCGGCTTCGTAACCTCAGTGGAGCGGTACGCGCGTGTGGAGGCACT 516
QY      508 GAGTACCCACTGAGCTGCTCCACGCGACGATTCATCACCATTGAAAGTCTCCGCG 567
Db      517 GTCTAGCAGCTGAGATGCGCCACACCCCGTTTCA-----CCGCA 558
QY      568 CATCGCGCTCGAAGCTCAGTAACGACCGGACAGAGACGAAAGAGTGAAGAGCT 627
Db      559 GCAAGCGCCGCTGCTTGTGAAGACGACGCGGTGACGCTGATGTGAGCGGCT 618
QY      628 TTATTCGAGCGCGAGCGCTCTCATGCGACTTCACCAAGAAATTCGAGATCAATTC 687
Db      619 CTACACCAAGGCGGAGTCTTACCTCCAGCTGGAGCGCCAGTTCAAGAGATCATCG 678
QY      688 GCACACCCCTGTTTTCGCGCTCTTGCACAGCGGTAAGATGCTCAACGATCTTCG 747
Db      679 CCACACCTCTGCTCTACACAGCTCGCGAGGAATACAGGCGCAACGAGCTTCACGA 738
QY      748 CCGTCTCCGTTGGATGACCGGTTGAAGACGCGCGAATACGTCGATGCGACTC 807
Db      739 GATTCACATCGGCGGAGACCGCGG-----AGCGGACCTTCGTGATGAGCGCTC 789
QY      808 AGCAGAAATCTTTTCCACTATCTACACGATACACAGCAAGAAAGCTTTTACCTT 867
Db      790 GCGCAACTGCTTT-----CGACTTCAGAACAGGCGGCAACGAGCGAGCC 834
QY      868 GGTGACGAACATCGCTGACACGACTGCGGCTTACGGCGGCTATGAGAAAGATTTGG 927
Db      835 GCGGAGAGCGCTTCAACTCTTCCCGCGGTGCTGAGCGCGCTGTCGCAACGC 894

```

QY	928	CGCTGCGCAGGTACAGAACTACTGCGCCACCAGCAAGTAATCTTAGTTGCGAGCTGAGACGCTA	987
Db	895	GTTCAGATCGGGATGCGAGAGCTTCATCCATCCAGCACTCATCTCTGCGCGGAGACCGGCTTGA	954
QY	988	TATCATATGGCAGAGGTATATATGTAATCTGGCGTCGGGAGCGCATTCGGCAACCCACAGATTCTCTA	1047
Db	955	AATCAAGGCTGACGCTACGTCAGTCTCCACCGCGGGCGGTCCACAACACAGCAGCTTCTCGT	1014
QY	1048	TAACTCGGGGCTT-----CTCTGGGGCTTACAGTCCAGGTCACGGCAGCAGCATGACTCGTTATATCC	1101
Db	1015	GAACCTCTGGCTTTTGGACAGCTATGCGCGGACCCGAAACCCCAAAACCCACCGGAGCTGCTCC	1074
QY	1102	CAACCTGGGGAGGTACATTCACGGAGCAGCCGATGGCACTTTTGGCAGATAGTCTTGGAGCA	1161
Db	1075	GTCCCTGGGGAGCTCATCTACCGAGCACTCGCTGCTTCTTGGCCAGACCGTGATGAGCAC	1134
QY	1152	GGAATTCGTCACACCGCT-----	1179
Db	1135	CGAGCTCATCGACACGCTGTAAGTCCGACATGCATCAGGGGAACCCCTGGCGAGCTGAC	1194
QY	1180	-----GGCGACGATCTTATGGA-CTGCCATGGTGGAAAGA	1215
Db	1195	GTACAGCTGACGTACAGCCCGGGCGGCTTCGACCAACAGCACCGGAGCTGTTGGAAGCA	1255
QY	1216	AGCGCTGCTCAACATATTTGCCAGAAACCCGACAGATGCTACTGCCATTCGTTCCGGA	1275
Db	1255	GAAGTGTAAAAAACCCATGATGACAGCACCAAGAGAGACCCGCTCCGATCCGTTGAGAGA	1314
QY	1276	TCCGGAACCCAGGTAAACAACCCCTTTACAGAAAGAACACCCCTTGGCACACGAGATTCA	1333
Db	1315	CCCCGAGCGCAGGTGTACCACTCTGTTCCAAACCGTCGACCCGTGGCACACTGCATCAATCA	1374
QY	1336	CCGCGATGCTTTTGTGACGGTGCCTGGTCTCTAGAGTGAATCTCTGTCATCATGCTGA	1395
Db	1375	CCGGATCTCTTCACTTACGGGCGCAGTGCAGCAAAAGCATTCGACTGCGTCTCATCTGTGA	1433
QY	1386	CTGCGCTGTGTTGGCGCAACCGACCTTGAAAGCAAAACCTTTTGTTTCCAGAACGA	1455
Db	1435	CTGGCGCTTCTCGCGCCGAGCGAGCCCAAGGAGAGAAACAGCTCTGTCTCGACAA	1494
QY	1456	TGTTCAAGACGGGTACGATATGCCGACCGACGTTCCAGTATATG---ACCAGCACTGC	1512
Db	1495	GATCACGAGACGCTACACATGCGCCAGCCACAGTTCGACTTCCGCTTCCGGCGGCGCG	1554
QY	1513	GTCAAACTGAGAGCAAGAAATGATGCCGATATGTGCGAAGTGGCGAGCACTTGGG	1572
Db	1555	CACGAGCAAGAGGGGAGAGCATGATGACCGACATGTCGTTATGTGGGGAAGATGG	1614
QY	1573	AGGTATTTGGCCCAAGTCCCGCCGCGCACTTATGATCCAGGCTTGCACCTTCACTTTCG	1632
Db	1615	TGGCTCTTACCCGCTCCCTCCGCCAATTCATGAGAGCTGCTGTCTTCTCACTCTGG	1674
QY	1633	GGGCACTACTGCAATTTGGCTTCGAC-----AAGCAACTACAGTGGCTGATTAACAATTC	1686
Db	1675	TGTTACGCAACCCAGGCGCTTCGAGAAAGAGAGACATCTGCTGCTCAACAGGACTC	1734
QY	1687	GCTGTCCTGGAGCTTGGCCAATCTTTATGTTGCAAGCAATGGACACATCAGAGCGGCTT	1746
Db	1735	GCGGCTGTTCCGGCTTCAAGAACCTCTCTCGTGGTGGCGGAAACATTCCTACGCGCTA	1794
QY	1747	CGGGGAGAAACCGACACTTACGTGATGTGGCACGCTATCAAGAGCGGAGAGACATCAT	1806
Db	1795	CGGGCGCAACCCGAGGCTCACCGGCAATGCTGCTCGCATCAAGAGTGGCGAGTCAATCAA	1854
QY	1807	CAATACACTCAAG	1819
Db	1855	GCAGAACTTCAAG	1867

DEFINITION	Sequence 1 from patent US 6146865.
ACCESSION	AR141573
VERSION	AR141573.1 GI:15101089
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 1869)
TITLE	Christensen,S., Lassen,S.,Plensted, and Schneider,P.
JOURNAL	Nucleic acids encoding polypeptides having pyranose oxidase activity
FEATURES	Patent: US 6146865-A 1 14-NOV-2000;
source	Location/Qualifiers
BASE COUNT	1.1869
ORIGIN	/organism="unknown"
	397 a 622 c 548 g 302 t
Query Match	10.1%; Score 212.4; DB 6; Length 1869;
Best Local Similarity	51.5%; Pred No.2.4e-33;
Matches	841; Conservative 0; Mismatches 656; Indels 135; Gaps 10;
269	GGCTACCAACAGAGATGAATGCAAGTTCAGAAAGATATTGACCGCTTCATAGTA 328
259	GGCTCACACAGAGAGACACCGCTGAGTACAGAGAGAACATGACAGAAATTCGTAATAGT 318
329	ATCAAGGAGAGCTTACACAGTCTCTGTTCTGTGCAGAAACAGACAGTGCCTACACTT 388
319	ATCAAGGGCAACTTATATGCCCTCTCGGGCCGCTCAACACGATGGTGTGACACGCTA 378
389	GATCCCGGAGCCTTGAGAGCGCGCCCTCGAAGTTACAGCATATTCAGCGGTAAATTCCT 448
379	ACCCCGGCTCATGGCAAGCTTCGACG-----TCTTCGTCCGCAACGGCGGCAATCCA 432
449	CACCAAGGGGAATTCGAGACTTGTCTGCGAGAGCCGTTACGCTGAGAGTGGCGGATG 508
433	GAGCAAGAGCCCGCTCGCAACCTTATGTGGCCAGGGGCGTACACCGCGTGTGCGCGCATG 492
509	AGTACCACCTGGAGGTGCTCCAGGACAGGATTCATCCACCCATGGAAGTGTCCGGGGC 568
493	TCTACGCACTGACGTGGCCGACGCGCGCTTCGAGAACTGCAG----- 537
569	ATCGGCCGCTCGAAGCTCAGTAAACGACCGGACGAGACGACGACAAAGATGGAACAGCTT 628
538	---CGCCGCTGCTGTGAAGAGACATCTCAAGGGGAGACGACCGGAGGGACAGGCTC 594
629	TATTCGAGAGCCGACGCTCTCATTCGGAGATTCCACCAAGAAATTGACGAGTCAATTGGG 688
595	TACAAAGAGGCCGAGTCTGACTTCAAGACCGGACGACCCCAATTGCGGAGTGCATCCGC 654
689	CACACCCCTTGTCTCGCGCTCTTTCGAAACAGCGGTCAAGAGATCTCTCAACGTTATTCGC 748
655	CACAACCTGCTGCTCAAGAAGCTGAGGAGGATCAAGG---GGGTGGGACATTTCCAG 711
749	CCTTCGCCGTTGGCATGCCACCGGTTGAAGAACGCGCGGGAATACGTGATGCGATCA 808
712	CAGATCCCGCTGCGCGGACGCGCCAGA-----GCCGAGCTTGTGAGTGGAGGCTCG 765
809	CGAGAAATCTTTTCCACTCTATCTACACGATGACAGACGAAAGAACTCTTTACCTCG 868
766	GGCACACCGTGT-----CGATTCGAGAACCGGCCGCAACAGAACGACCGC 810
869	CTGACGAACCATCGTCACACGACTGGCGCTTACGGCGGGTATGAGAAAGAAATGGC 928
811	CCGAAGCAGCGCTTCAACCTCTCCCGCCGCTCGCGTGCACAGAACGTGAGCGCGCATAC 870
929	GCTGCGGAGGTCAGGAATCTATGCGCCACACGAGATCTTAATTGACACTGGAACAAT 988
871	CGGAATCTGGAGATCTGAGGCTGGATGTCGCGACCTCCACGGGGGCAAGACATCAC 930
989	ATCATGGCGAAGGTATATGACTGGCGCTCGGAGAGATCGGCAACCCACAGATTTCTAT 1048
931	ATCAAGGCAAGGATATCATCTTCACGCGGGGCGGGTTCACAAAGCGGACGCTCTCGGG 990

QY 1049 AACTGGGCTT-----CTTGGGCTACAGGTACAGCCAGCAATGAC--TCGTTGATC 1099
 Db 991 GCCTTGATTCGGGAGCTGGGTGCTCCGAGACCCGCAAGCCGCTGCGCTCTGCTG 1050
 QY 1100 CCCAACCTGGGAGGTACATACAGGAGCAGCCGATGGCATTTTGGCCAGATGCTTGAGG 1159
 Db 1051 CCGTACTGTTGGGAGCCCATCATCACAGAGACGCTGCTTCTGCGACAGCCGTCATGAGC 1110
 QY 1160 CAGATATTCGTGACAGCTGGCGCAGATC----- 1190
 Db 1111 AGGAGCTCATCAAGGTGACAGCGGATATGACATTTGTGGCAAGCCCGCCACCG 1170
 QY 1191 -----CTTANGACTGCCATGTGGA 1213
 Db 1171 GACTATAGCTGACGTATACCCCGGAGCAACCCGACACAGACAGCAGCCGAGCTGTGGAAC 1230
 QY 1214 GAAGCCGTGCTCACAATATTTGCCAAGAACCCGACAGATGCACTGCCCATTCGTTCCGC 1273
 Db 1231 GAGAAAGTGAAGAGACATGATGACACAGAGAGACCCGCTCCGATCCCGTTGAG 1290
 QY 1274 GATCCGGAACCCAGGTAAACAACCCCATTTACAGAAACACCCTGGACACAGCATTT 1333
 Db 1291 GACCCGAGCCGAGGTACACAGCTGTTTCAAGGACAGCACCACCCAGCAGCCAGATT 1350
 QY 1334 CACCCGATGCTTTTCTGACGGTCCGTCGGTCTGAGGTGACCTCTGTCATGCTC 1393
 Db 1351 CACCCGAGCCCTTACGTACGGCGCCGTGACAGAGATGACATCGCGGCTCATGCTC 1410
 QY 1394 GACCTGCTGTTGGTGGCAACCCGACCCCTGAAGCAAAACCTTTGTTTCCGAAAC 1453
 Db 1411 GACTGCGCTTCTTCCGAGCAGCAGCCAGCCAGAGAGAGAAAGCTATGTTCTGAGC 1470
 QY 1454 GATGTTCAAGAGCGGTACAGTATGCCGAGCCGACGTTGATGATGACCCAGCATGCG 1513
 Db 1471 AAGATCAGGAGCGGTACAAACCTCCGAGCCGAGCTTGACTCGGCTCCCGGGG-- 1528
 QY 1514 TCAAAGTGAAGCAAGAAATGATGCCGATATGTCGAACTGGCGAGCAACTTGGGA 1573
 Db 1529 ----GCCGGAAGCGAGAGACATGATGATGCGCATGTGTCATGTGCGGAGAAATCGGT 1584
 QY 1574 GGTATTTTCCAGCGTCCCGGAGTTTATGATGATGCAAGCCCTTCACCTTCACTCTGGG 1633
 Db 1585 GATTTCTGCTGCTGGGTCTACCCACAGCTTCATGAGACCCGCTTCTTCTGCACTTGGT 1644
 QY 1634 GGGACTACTCGCATGTTGGCTTGACAA-----GGCAACTACAGTGGCTATAACAACCTG 1687
 Db 1645 GGGAGCAGCCGATGGGCTTCGAGAGAGGGGAGCAAGTGTGGCTTCAGACCCGATCA 1704
 QY 1688 CTGGTCTGGGACTTTGGCAATCTTTATGTTGAGGCAATGCAATCAGACAGGGGCTTC 1747
 Db 1705 CCGTCTTCTGGCTTCAAGAACTCTTCTCGGCGGCTGCGGAACTCCCAACCGGCTAC 1764
 QY 1748 GGGGGAACCCGACCTTACGTGATGATGTCACAGCAGTATCAAGAGCCGAGAGCATATC 1807
 Db 1765 GCGGGAACCCGACGCTACCGCAATGCTGCTTGGCATCAAGAGTGAATCAAG 1824
 QY 1808 AATACACTCAAG 1819
 Db 1825 AAGAACTTCGAG 1836
 RESULT 5
 LOCUS E11766 1869 bp RNA linear PAT 29-SEP-1997
 DEFINITION cdna of pyranose oxidase.
 ACCESSION E11766
 VERSION E11766.1 GI:22025402
 KEYWORDS JP 1996205861-A/1.
 SOURCE unclassified
 ORGANISM unclassified
 REFERENCE 1 (bases 1 to 1869)

AUTHORS Nishimura, I., Okada, K., Nanbara, T., Kawai, G., Koyama, T. and Suzuki, M.
 TITLE NEW PYRANOSE OXIDASE, PYRANOSE OXIDASE GENE, NEW RECOMBINANT DNA AND PRODUCTION OF PYRANOSE OXIDASE
 JOURNAL Patent: JP 1996205861-A 1 13-AUG-1996;
 KIKKOMAN CORP
 COMMENT OS Coriolus versicolor
 PN JP 1996205861-A/1
 PD 13-AUG-1996
 PE 24-MAY-1995 JP 1995124835
 PR 07-DEC-1994 JP 94P 304086
 PI NISHIMURA IKUKO, OKADA KIMIHARU, NANBARA TOMOYUKI, PI KAWAI GENSHIRO,
 KOYAMA TAJIJI, SUZUKI MASARU
 PC C12N9/04, C12N15/09, (C12N9/04, C12R1:19), (C12N15/09, C12R1:01);
 CC strandedness: Double;
 CC topology: Linear;
 FH key Location/Qualifiers
 FT source 1..1869
 FT /strain='ps4a', /organism='Coriolus versicolor' FT
 FEATURES
 source 1..1869
 /organism='unidentified'
 /mol_type='genomic RNA'
 /db_xref='taxon:32644'
 BASE COUNT 408 a 593 c 529 g 339 t
 ORIGIN
 Query Match 10.0%; Score 210.2; DB 6; Length 1869;
 Best Local Similarity 50.4%; Pred. No. 6; 9e-33;
 Matches 857; Conservative 0; Mismatches 713; Indels 129; Gaps 9;
 QY 200 GAGATGGAGCTGCTGATAGCTTCTACGCTGTTAATGCCGAAGAAAGCACTGAGTTCC 259
 Db 190 GACCTGCTGAGAGCCGCTTACAAAGTCCGATGTTGACATCGGGAAATTTGACTTGGC 249
 QY 260 TACGTTCTGGGTACCAAGAAGATGAATGCAATGCAATGCAATGCAATGCAATGCAATG 319
 Db 250 CTGAAGATCGTGGTCCCAAGAAGAACACCGCTGATATACAGAAAGCAATTTGACAGTTT 309
 QY 320 GTCAATGTAATCAAGAGGACCTTACAAAGTCTGTTCTGTTCTGCAAAACCAAGAGTG 379
 Db 310 GTGAAGCTATTCAGGAGGCAATGATGTTGTTCCGTTCCGTTCAATACCTCGTATC 369
 QY 380 CCTACCTGATGCCGAGGAGCTGGAGCGGCCCTCGAAGTTCAACCATATGAGAGGT 439
 Db 370 GACAGCTCAGCCCGAGCTTGTGGCAAGCTTATCG-----TTCTTCTCGCAATGAGC 423
 QY 440 AAAAATCTTCACAGCGGGAATTCGAACTTGTCTGCGAGGCGCTTAACGCGTGAATC 499
 Db 424 TCGAACCAGAGAGAGACCCGCTTCTGTAACCTAGTGTGCAAGCGGTTCACGCGTGTGTC 483
 QY 500 GCGGCAATGATACCACTGAGAGTCTTCAGCCAGCAATTCATCACCATGAGAACT 559
 Db 484 GGAGCATGTCCACGACTGACATCGGAGACACCCGCTTTGA----- 527
 QY 560 CTCGCGGATGAGCGGCTCCGAAGCTCAATACAGCCGAGAGAGAGCAAGAAAGCTGG 619
 Db 528 --CCGAGAGAGCGGCGCTTGTCTGGAAGAGAGACCAAGAGAGCTGACGCGCGAGTGG 585
 QY 620 AACGACTTATTCAGAGCGAGCGCTTCATCGGACTTCCACCAAGAAATTCAGAGAG 679
 Db 586 GACCGCTTACCAAGAGCGCGAGCTATCTTAAGAGACCGGAGAGAGCAAGTTCAAGAG 645
 QY 680 TCAATTCGACACCCCTTGTCTGCTTTGCAAGAGCGGTACAGAGATGTCACCT 739
 Db 646 TCGATCCGCAACAACCTGCTCAACAAGCTCGCGAGAGAAATACAAAGTCAAGCGAGC 705
 QY 740 ATCTTTGCGCTCTCCGTTGGCATGCAACCGGTTGAAGACGCGCGGGAATATCGTGA 799
 Db 706 TTCCAGCAGATCCCGCTCGGAGCAAGCGCTG-----CAGTCCGACCTTCGTCGAG 756

Db 215 GAGTCGTGAGAGCGGTTACAAAGTCCGATGTCACATCGGGAAATTTGACTCTGGC 274
 QY 260 TACGTTCTGCTACCAAGAGATGAATGAGTTCCAGAAAGATTTAGACCGCTTC 319
 Db 275 CTGAAGATCGGTGCCCCCAAGAAAGACCGCTGCAATACAGAAACATTGACAACTTT 334
 QY 320 GTCAATGTATCAAGAGAGGACCTTACAAAGTCTCTGCTTCCCTGTCAAGAAACAGAACGTG 379
 Db 335 GTGAAGGTATCGAGGGCCAAATTGATGTCTGTTCCGTTCCCGTCATACCTCTGTATC 394
 QY 380 CTTACACTTATATCCCGAGGCTTGAGCGGCCCTCCGGAAGTTCCAGCCATATGCAAGCT 439
 Db 395 GACACGCTCAGCCGACGCTTGTGCAAGCTTTCATCG-----TTCCTTCGTCGCAAAAGGC 448
 QY 440 AAAATTCCTCACCAGCGGGAATTCGAGACTTGTCTCGAGAGCGCGTCAAGCGCGTGGATC 499
 Db 449 TCGAACCAGAGAGAGAGCCGCTTCGTACCTCAGTGTGTAAGCGGTGACGCGTGTGTC 508
 QY 500 GGGGCGATGATGCCACCTGAGCGTGTCCACGCGCAGGATTCATCCACCCATGGAAGT 559
 Db 509 GAGGCGATGTCACGCGACTGACATGCGGACACCGCGCTTGA----- 552
 QY 560 CTCGCGGCGATCGGCGCTCGAAGCTAGTAAAGACCGCGCAGAGACGACAAAGAGTGG 619
 Db 553 --CCGGAGAGCGCGCGCTTCTGTGAAGAGACGACGCGTGTGACGCGCGAGTGG 610
 QY 620 AACGAGCTTATTCCGAGCGGAGCGCTCTATCGGGAGCTTCCACCAAGATTCGAGAG 679
 Db 611 GACCGGCTGTACACCAAGGCGCGAGTCACTTCAAGCGCGGAGCGACCGCTCAAGAG 670
 QY 680 TCAATTCGACACCCCTTGTCTGCGCTTTTGCAMAGACGCTACAAAGATTCGTAAGT 739
 Db 671 TCGATCGGCGACAACTCGTGTCAACAGACTCGCGAGAGATTAAGAAAGTCAGCGGAC 730
 QY 740 ATCTTTCGCGCTCTCCGCTTGAGTGCACCGGTTGAAGAGCGCGGAGTAATACGTGAA 799
 Db 731 TTCCAGAGATCCCGCTCGCGGCAACGCGTCG-----CAGTCCGACCTTCGTGAG 781
 QY 800 TGGCACTCAGCAGAAATTTTTCACACTATCTACAAAGATGCAAGAGAAAGACTC 859
 Db 782 TGGAGCTCGGCGAAGACCGTGT-----CGACCTCCAGAAAGCGCGAAGC 826
 QY 860 TTTACCTGCTGACGAACCATCTGTCACAGACTGCGCTTACGGGCGGGTATGAGAG 919
 Db 827 ACGGAGCGCGCAATGAGCGCTTCAACCTCTCCCGGTGTGATGTGAGCGCGTGTG 886
 QY 920 AAGATTGGCGCTGCGGAGTCAAGATCTACTGCGCACGAGAACTCTAGTTGCGACGTG 979
 Db 887 CGCAACACGTGAACTCCGAGATCGAGAGTCTGCACATCCACGACCTCATCTCGGGGAC 946
 QY 980 GACAGCTATATCAATGCGGAAGTATATGTAAGTGGCGTGGGAGGAGTGGCAACCCACAG 1039
 Db 947 CGCTTGAATCAAGAGCAGACGTGTCTTACACGCGGGGCGGTCCACACGCGCAG 1006
 QY 1040 ATTCTCTATATCTGCGGCTT-----CTCTGGGCTACAGGTCAAGCCACCAATGACTCG 1093
 Db 1007 CTCTCGTGAATCTGCTTGTGACAGCTGCGCGCGCGGACCCCGGAGACCCGCGCAG 1066
 QY 1094 TTATATCCCAACTCGGGAGGTATACGAGAGACGCGATGTCATTTTGGCAGATATGTC 1153
 Db 1067 TTCTCGCGCTGCTGGGAACCTACATCCAGAGAGTCTGCTCTTCTCCAGAACGCTG 1126
 QY 1154 TTGAGGAGGAATTCGTGACAGCGTG-----GGCGAGATTCCTTATGAGCTGCG----- 1203
 Db 1127 ATGAGCAGCCGAGCTATCGAGGCTCAAGTCCGACATGATCATAGGGCAACCCCTGGC 1186
 QY 1204 -----ATGG 1207
 Db 1187 GATCTGGGGTACAGCGTACAGCGCCCGGGCGGAGACCAAGCAAGCAGCGGAGTTGG 1246
 QY 1208 TGGAAAGAAAGCGCTTGTCTCAATATTTGCCAAGAACCCGACAGATGACCTGCCATTCG 1267
 Db 1247 TGGAGGAAAGGTGAAGAACACATGATGACAGCACCGAGGAGACCCGCTTCCAAATCCG 1306

QY 1268 TTCCGCGATCGGAAGCCCAAGTAAACCCCATTTTACAGAAAGAACACCCCTGGACACG 1327
 Db 1307 TTGAGAGACCCCGAGCGCGAGGTACACCTTGTTCAGGACATGACCCGCTGGACACT 1366
 QY 1328 CAGATTCACCGCGATGCTTTTTCGTAAGGTGCGCTGCTGCTGAGTGGAGTCTGCTG 1387
 Db 1367 CAGATTCACCGCGATGCTTTTTCGTAAGGTGCGCTGCTGAGTGGAGTGGAGTCTGCTG 1426
 QY 1388 ATGCTGACCTGCGCTGCTTTTGGCGCAACCGACCTTAAGCAAAACCTTTTGGTTTC 1447
 Db 1427 ATGCTGACCTGCGCTTCTTCCGCGGAGGAGGCAAGGAGGAAACAGCTCTGCTTC 1486
 QY 1448 CAGAGATGATTCAGAGGAGGAGTACAGATGCGGAGCGGAGCTTCAATATGCAATGCA 1507
 Db 1487 TCGGACAAATTTACGACACGTACACATGCTCCGAGCGGAGCTTCCGCTTCCG 1546
 QY 1508 ACTG--CGTCAACGTAAGACGAAGAAATGATGGCGATATGTGCAAGTGGCGAGC 1564
 Db 1547 GCGGCGCGACAGACGAAGGAGGAGGAGATGATGATGATGATGATGATGATGATG 1606
 QY 1565 AACTTGGAGATTTTGGCCAGCTCCCGCGAGTTTATGATTCAGGCTTGGACTT 1624
 Db 1607 AAGATTGTGTGCTTCTGCGCGGCTCCCTCCGCAATTCATGAGCGCGCTTGTCTT 1666
 QY 1625 CATCTGCGGAGTACTTCCGATTTGCTTGCAC-----AAGGCACTACAGTGGCTGAT 1678
 Db 1667 CACTCGGTGTGTACGACCGCATGGCTTGGACGACGAGAGGACAAAGTGTGCTCAAC 1726
 QY 1679 AACAACTGCGTGTCTGAGGACTTGCATCTTTATGTTTGCAGGCAATGAGCAACATCAG 1738
 Db 1727 ACGGACTGCGGCGGCTTGTGCTCAAGAACTGTCTCGGTGCTGCGGAAACTTCCC 1786
 QY 1739 ACGGCTTGGCGGAGAACCCGACACTTACGTGATGTGCCAGCTATCAAGAGCGGAG 1798
 Db 1787 ACGGCTGAGCGGCGGAGACCGACGCTACCGCAATGTGCTGCGATCAAGAGTGGAG 1846
 QY 1799 AGCATCATATACATCA 1817
 Db 1847 TACATCAAGAACTTCA 1865

RESULT 8
 E32737
 LOCUS E32737 1902 bp DNA linear PAT 18-JUN-2001
 DEFINITION Panus cornucopiae-derived antitumor protein and gene thereof.
 ACCESSION E32737
 VERSION E32737.1 GI:13017467
 KEYWORDS JP 1999315096-A/1.
 SOURCE Pleurotus cornucopiae (cornucopia mushroom)
 ORGANISM
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 Agaricales; Pleurotaceae; Pleurotus.
 REFERENCE 1 (bases 1 to 1902)
 AUTHORS Akihiro, M., Koji, I., Kazuaki, K., Tomohide, S. and Ryoko, A.
 TITLE Panus cornucopiae-derived antitumor protein and gene thereof
 JOURNAL JP 1999315096-A 1 16-NOV-1999;
 NYUFDO KURIKISHON GIUTSUENKUYUMIAI
 OS Pleurotus cornucopiae
 PN JP 1999315096-A/1
 PD 16-NOV-1999
 PF 07-AUG-1998 JP 1998236349
 PR
 PI AKIHIRO MORITA, KOJI IZUMO, KAZUAKI KIDO, TOMOHIDE SAKA, PI
 RYOKO AOKI
 PC C07K14/375,A61K35/84,A61K38/00,C07K16/14
 CC
 FH Key Location/Qualifiers
 FT source 1..1902
 FT Location/Qualifiers
 1..1902
 /organism="Pleurotus cornucopiae"
 /organism="Pleurotus cornucopiae"

TITLE Novel protein, gene encoding the same and method for utilization
JOURNAL Patent: JP 2000083675-A 1 28-MAR-2000;
JAPAN TOBACCO INC., SOCIETY FOR TECHNO-INNOVATION OF AGRICULTURE
FORESTRY AND FISHERIES
COMMENT OS Tricholoma matsutake
PN JP 2000083675-A/1
PD 28-MAR-2000
PE 08-SEP-1998 JP 1998270606
PR YOSHIMITSU TAKAKURA, SHIGERU KUMATA, SHOZO OHTA
PI C12N15/09,A01N37/12,A01N63/00,A01N63/02,A01N65/00,C07K1/14, PC
C07K1/18,
PC C07K1/30,C07K14/375,C12N1/19,C12N1/21,C12N5/10,C12N5/10, PC
C12P21/02,
PC C12Q1/68//(C12N15/09,C12R1:645),C12N15/00,C12N5/00,C12N5/00,
PC (C12N15/00,C12R1:645)
CC
FH Key Location/Qualifiers
FT source 1. 1946
/organism="Tricholoma matsutake".
FEATURES
source 1. 1946
Location/Qualifiers
/organism="Tricholoma matsutake"
/mol_type="genomic DNA"
/db_xref="taxon:40145"
BASE COUNT 484 a 516 c 475 g 471 t
ORIGIN
Query Match 6.1%; Score 129; DB 6; Length 1946;
Best Local Similarity 52.5%; Pred. No. 3.3e-16;
Matches 387; Conservative 0; Mismatches 305; Indels 45; Gaps 3;
QY 1106 CTGGGAGGATGCTACGAGCGGAGGATTTGGCAGATGCTTGAGCGGAA 1165
DB 1064 CTTGGTCGCTACTCTGACGACGATCCATGACTTTTGTGAGATGTTCTCAAGAGAGC 1123
QY 1166 TTGCTGACAGCGCTGGCGGACGATCCCTTATGAGACTGCCATGTGGAAGAAGCGTTGCT 1225
DB 1124 ATAGTGATGATTCATCGCTACTGACCT-----CGCTTCGCTGGGAAGGTTGAG 1171
QY 1226 CAACATATTGCCAAGAACCCGACAGATGCATGCCATTCGTTCCGCGATCCGGAACC 1285
DB 1172 GCGCACAAGAGAAGACCCCGATGACGTGCTGCCATTCATCCACAGAGCTGAACCT 1231
QY 1286 CAGGTAAACACCCATTTACAGAGAACAACCCCTGGACACGAGATTCACCGCGATGCT 1345
DB 1232 CAGATGATGATTCGTAACGCTGACGCTTCCCTTGATGATGTTCAAGTCCATTCATGCA 1291
QY 1346 TTTTCGTAGCGTCCGCTCGTCTGAGTGGAGCTCTGCTCATGCTGACCTGCGCTGG 1405
DB 1292 TTT-----GCTGATTTGGACCCCAAGCCGCGCTGTGCTCGATCTGAGGTTT 1345
QY 1406 TTTGGCGCAACCGACCTGAGCAAAACCTTTTGGTTTCC----- 1448
DB 1346 TTGCGCAAAATCAGATTTGTCGAAGAAATCGAGTACCTTCGCTCGAACCCTAAGCTA 1405
QY 1449 -----AGAAAGATGTTCAAGCGGGTACAGATGCCGAGCGAGCTCAAGATAT 1498
DB 1406 CCGGACTGGGAAGCGGGTGTACAGACACTTATGATGACCAAGCCGAGATTCATGTC 1465
QY 1499 CGACCCAGACATGCTCAAAACGTGAGAGCAAGGAATAATGATGCGATATGTCGAAGTG 1558
DB 1466 AAGCGGACCAAGCCGATGGAAGCGTGAACCAAGATGATGATGATGATGACCAACGCTC 1525
QY 1559 GCGAGCAACTGGGAGGTTATTTGGCCAGCTCCCCCGAGTTTATGATTCAGGCGCTT 1618
DB 1526 GCGAATACATCTGGCGCGGTACCTTCCTGGCTCTACCTCAATTTATGGAACCTGGCTC 1585
QY 1619 GCACCTCATCTGGGAGCTACTGCAATTTGCTTGAAGCAAGCACTACAGGGGCTGAT 1678
DB 1586 GCACAGCATATCCGGAAGCTACTCGATCGGAGCAAGATGATCAAACTTGTGTGCTGAT 1645
QY 1679 AACCACTCGCTGGTCTGGGACTTTGCCAATCTTATGTTGACAGCAATGACCAATCAGG 1738

DB 1646 CCACATCAAAAGGTCATATACTTCGACAACTCTGGCTGGCGGGAATGGGTCAATCCA 1705
QY 1739 ACGGCTTGCGGCGAGAACCCGACACTTACTGATGTGGCCACCTATCAAGACCGGAGG 1798
DB 1706 GATGCACTGCTGCAACCCGACTCGTACGAGCGCTCGGTATCGCTTAAGGCTGTGAG 1765
QY 1799 AGCATCATCAATACACT 1815
DB 1766 GCTGTAGTCACTTACT 1782
RESULT 11
E37379 1946 bp DNA linear PAT 31-JAN-2002
LOCUS E37379
DEFINITION Reagent for assaying pyranose containing pyranose oxidase.
ACCESSION E37379
VERSION E37379.1 GI:18626680
KEYWORDS JP 2000175698-A/1.
SOURCE Tricholoma matsutake
ORGANISM Tricholoma matsutake
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Tricholomataceae; Tricholoma.
REFERENCE 1 (bases 1 to 1946)
AUTHORS Takakura,Y., Kuwata,S., Ohta,S. and Usami,S.
TITLE Reagent for assaying pyranose containing pyranose oxidase
JOURNAL Patent: JP 2000175698-A 1 27-JUN-2000;
JAPAN TOBACCO INC
OS Tricholoma matsutake (pine mushroom)
PN JP 2000175698-A/1
PD 27-JUN-2000
PE 16-DEC-1998 JP 1998357423
COMMENT
QY 1106 CTGGGAGGATGCTACGAGCGGAGGATTTGGCAGATGCTTGAGCGGAA 1165
DB 1064 CTTGGTCGCTACTCTGACGACGATCCATGACTTTTGTGAGATGTTCTCAAGAGAGC 1123
QY 1166 TTGCTGACAGCGCTGGCGGACGATCCCTTATGAGACTGCCATGTGGAAGAAGCGTTGCT 1225
DB 1124 ATAGTGATGATTCATCGCTACTGACCT-----CGCTTCGCTGGGAAGGTTGAG 1171
QY 1226 CAACATATTGCCAAGAACCCGACAGATGCATGCCATTCGTTCCGCGATCCGGAACC 1285
DB 1172 GCGCACAAGAGAAGACCCCGATGACGTGCTGCCATTCATCCACAGAGCTGAACCT 1231
QY 1286 CAGGTAAACACCCATTTACAGAGAACAACCCCTGGACACGAGATTCACCGCGATGCT 1345
DB 1232 CAGATGATGATTCGTAACGCTGACGCTTCCCTTGATGATGTTCAAGTCCATTCATGCA 1291
QY 1346 TTTTCGTAGCGTCCGCTCGTCTGAGTGGAGCTCTGCTCATGCTGACCTGCGCTGG 1405
DB 1292 TTT-----GCTGATTTGGACCCCAAGCCGCGCTGTGCTCGATCTGAGGTTT 1345
QY 1406 TTTGGCGCAACCGACCTGAGCAAAACCTTTTGGTTTCC----- 1448

Db 1346 TTGGCAATGATATTGCGAAGAAATGAGTACTTTCGTCGCAACCTTAAGCTA 1405
OY 1449 -----AGAACGATGTTCAAGACGGGTACTATGCGCAGCGAGCTTCAGATAT 1498
Db 1406 CGGACACTGGGAAGCGGGTTCACAGACACTATGGAATGCCAGCCGACATTCATGTC 1465
OY 1439 CGACCACTGCTGCTCAACGCTGAGAGCAAGAAATGATGCGCATATGTCGAGATG 1558
Db 1466 AAGCGGACCAACGCCGATGAGACCTGACCAAGATGATGAATGATGACCAACGTC 1525
OY 1559 GCGAGCACTGGGAGGTATTTGGCCACGTCGCCCGCGAGTTATGATCCAGCCCTT 1618
Db 1526 GCGACATGACTGGGGGCTACCTTCCTGCTCCTCACTCAATTTATGCACTGCTGTC 1585
OY 1619 GCATCTATCTTGGGGGACTACTGCACTTGGCTTCGACAAAGCACTACAGCTGCTAT 1678
Db 1586 GCACAGCACTACGCGGAGAACTACTGCGAGACAGATGATCAAACTTCCTGCTGCTAT 1645
OY 1679 AACCACTGCTGCTGCTGAGACTTTGCCAATCTTTATGTTGCAAGCAATGCAACATCAGG 1738
Db 1646 CCGAAGCTAAAGGTTTCATGCTGACAACTGTGGGTCGGCGGAATGGGTGCAATCCA 1705
OY 1739 ACGGCTTCGGGAGAACCGGACCTTACGTCGATGTCGACGCTATCAAGAGCGCGAGG 1798
Db 1706 GATCGACTGCTTCGACACCGACTGCTGAGAGCGCTGATGCGCTTAAGAGTGTGAG 1765
OY 1799 AGCATCATCAATACACT 1815
Db 1766 GCTGTAGTCACTTACT 1782

RESULT 12
LOCUS AB043883 1946 bp mRNA linear PLN 29-NOV-2002
DEFINITION Tricholoma matsutake P20 mRNA for pyranose 2-oxidase, complete cds.
ACCESSION AB043883
VERSION AB043883.1 GI:25553432
KEYWORDS Tricholoma matsutake
SOURCE Tricholoma matsutake
ORGANISM Tricholoma matsutake
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Tricholomataceae; Tricholoma.
REFERENCE 1 Takakura, Y.
AUTHORS Purification, characterization and primary structure of a pyranose
Oxidase from Tricholoma matsutake
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1946)
AUTHORS Takakura, Y.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-2000) Yoshimitsu Takakura, JT Plant Innovation
Oxidase Center, 700 Higashibara Toyoda-cho, Iwata-gun, Shizuoka-ken
433-0802, Japan (E-mail: yoshimitsu.takakura@ims.jti.co.jp,
Tel: 81-538-33-8291, Fax: 81-538-33-6046)
FEATURES
Source
1. 1946
Location/Qualifiers
/organism="Tricholoma matsutake"
/mol_type="mRNA"
/db_xref="taxon:40145"
/note="synonym: Tricholoma nauseosum"
1. 1946
/gene="P20"
101. 1795
/gene="P20"
/EC_number="1.3.10"
/codon_start=1
/product="pyranose 2-oxidase"
/protein_id="BAC24805.1"
/db_xref="GI:25553433"
/translation="MPRLSRKINDLQSRGDLTSSODEIVHTDVFAGSGPIAC
TYARHIDNTSTKVVMAETGSDNPYVIAHHRSTKRFQDKEFVNICALQPIST
SPSDTYQPTLAIVAMAPPIDBAEQVIMGNPQEGNLDPGSAVTRIVGGMATHTMT

CACPTPDEERVENNEVDKQEFDALEAKTLLNWSDOYDLSHQIVYKFTLQGTLLDA
SRGVTTPLELGEERTDNDIYVTTGADVLDDVPRSPREVLVTEPRVKTIVSETNPT
OVVALLLENTLNDLIVVAVSFLVACAVCTPDIILMSNLRPHALGRLVLEOSMTFC
OIVLKRSIVDSIADPREFAKVEAHKHKHPDDVLPPEHEPOVMIPLYTSDPFMHVO
VHRAFGVGRKADPRVVDLRFEGKSDIYVENRYTEGPNKLDMEAGVDTYGMQO
PFFHVKRTNAGDDNDQRMNMTNANLGSITGLSGSYQFMAPRLAQHTITTTTATGTD
DOTSVADPTSKVHNEDNLVGNNGCTIPATCNPRISVAITALGAEAVSYLGVSS-
BASE COUNT 484 a 516 c 475 g 471 t
ORIGIN
Query Match 6.1%; Score 129; DB 8; Length 1946;
Best Local Similarity 52.5%; Pred No. 3.3e-16;
Matches 387; Conservative 0; Mismatches 305; Indels 45; Gaps 3;
OY 1106 CTGGGAGATACATACGAGACCGCGATGCGATTTTCCAGATAGCTTGAAGCAGAA 1165
Db 1064 CTGGTGCGTACCTCAGGCAACAGTCCATGACTTTTGTCAGATGTTCTCAAGAGAGC 1123
OY 1166 TTGTCGACACCGTGGCGGACGATCTTATGAGACTGCCATGTGGAAGAACCGTGTCT 1225
Db 1124 ATAGTCGATTCATCGCTACTGACCT-----CGCTTCGCTGCGAAGGTTGAG 1171
OY 1226 CAACATATTTGCCAAGAACCGGACAGATGACCTGCCATTCGCTCCGCGATCCGAAACC 1285
Db 1172 GCGCAGAGAAAGAACCGCGGATGACCTGCTGCCGATTCATTCACGAGACTGAACT 1231
OY 1286 CAGGTAAACACCCATTTTACAGAAACACCCCTTGGCAGACGATTTACCGCGATGCT 1345
Db 1232 CAAGTGATGATTCGCTTACACGTCGACTTCCTTGGCATGTTCAGAGCTCATCGGTATGCA 1291
OY 1346 TTTTCGTACGCTGCGCTGCTCTGAGTGTGACATCTCGTGTATCTGTGACCTGCGCTGG 1405
Db 1292 TTT-----GGTGTGTTTGGACCCAGGCGGCGCTGTGTGTGCTGATCTGAGGTTT 1345
OY 1406 TTTGGCGCAACCGACCCCTGAGCAAAACCACTTTGTGTTTCC----- 1448
Db 1346 TTGGCAATATGATATTGTTGGAAGAAATGAGTGAAGTTCGCTGCGCAACCTTAAGCTA 1405
OY 1449 -----AGAACGATGTTTCAAGACGGGTACAGATGCCGACCGGCTTCAGATAT 1498
Db 1406 GCGCACTGGAAAGGGGTGTTACAGACACTTATGGAATGCCAGACGACATTCATGCTC 1465
OY 1499 CGACCACTGCTGCTCAAAAGCTGAGAGCAAGAAATGATGCGCATATGTGGAAGTG 1558
Db 1466 AAGCGGACCAACGCCGATGAGACCTGACCAAGAGATGATGATATCAACAACTC 1525
OY 1559 GCGAGCACTTGGGAGGTATTTGGCCACGTCGCCCGCGAGTTATGATGATGAGGCTT 1618
Db 1526 GCGAACAATGCTGGGGGCTTACCTTCCTGCTGCTCACTCAATTTATGACACCTGCTC 1585
OY 1619 GCATCTATCTTGGGGGACTACTGCGATGCTTGCCTTGCACAAAGCACTACAGTGGCTGAT 1678
Db 1586 GCACAGCACTACGCGGAGAACTACTGCGATGCTGCGGACAGATGATCAAACTTGTGTCGAT 1645
OY 1679 AACCACTGCTGCTGCTGAGACTTTGCCAATCTTTATGTTGCAAGCAATGCGACATCAGG 1738
Db 1646 CCGAAGCTAAAGGTTTCATGCTGACAACTGTGGGTCGGCGGAATGGGTGATCCA 1705
OY 1739 ACGGCTTCGGGAGAACCGGACACTTACGTCGATGTCGACGCTATCAAGAGCGGAGG 1798
Db 1706 GATCGACTGCTTCGCAACCGCACTGTCGAGCGTTCGCTATGCGCTTAAGGCTGTCAG 1765
OY 1799 AGCATCATCAATACACT 1815
Db 1766 GCTGTAGTCACTTACT 1782

RESULT 13
LOCUS AB070201 2013 bp mRNA linear PRI 16-AUG-2001
DEFINITION Macaca fascicularis testis cDNA clone: Qtsa-17226, full insert
ACCESSION AB070201

[illegible]

RESULT 14					
BC049706	LOCUS	BC049706	589 bp	mRNA	linear
	DEFINITION	Mus musculus, clone IMAGE:6708084,			ROD 01-APR-2003
	ACCESSION	BC049706			
	VERSION	BC049706.1	GI:29436895		
	KEYWORDS				
	SOURCE				
	ORGANISM	Mus musculus (house mouse)			
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	AUTHORS	1 (bases 1 to 589)			
	TITLE	Strausberg,R.			
JOURNAL		Direct Submission			
		Submitted (31-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK		NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT		Contact: MGC help desk Email: gcgabs-remail.nih.gov Tissue Procurement: Dr. Jonathan Kuo, NIMH CDNA Library Preparation: Michael Brownstein Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@cgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chin, Chris Fjell, Erin Garland, Ran Guln, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carlie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Speedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.			
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 45 Row: n Column: 24 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.					
FEATURES	source	Location/Qualifiers			
		1..589			
		/organism="Mus musculus"			
		/mol_type="mRNA"			
		/db_xref="taxon:10090"			
		/clone="IMAGE:6708084"			
		/tissue_type="Testicle, mouse"			
		/clone_lib="NIF_MGC_169"			
		/lab_host="DHIOB"			
		/note="Vector: pONR-LIB"			
BASE COUNT		314 a 96 c 88 g 91 t			
ORIGIN					
	Query Match	4.5%: Score 94.4; DB 10: Length 589;			
	Best Local Similarity	80.9%: Pred. No. 3.9e+09;			
	Matches 110; Conservative 0; Mismatches 26; Indels 0; Gaps 0;				
Db	1971 TGTGATTACTGAAGCTGTACTTGTCGTGAATAGTTATGGCACAATCATTCATGTTAA	2030			
OY	II IIII II IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII	2090			
Dd	375 TGATTATTATTATTTCCCGAGATTATATGATAAAATTAAAGGCCATATGTTCAAAAAAAA	434			
OY	2031 AA	2090			
Dd	435 AA	494			
OY	2091 AAAAAAAAAAAAAAAAAAAAAA 2106				
Dd	495 AAAAAAAAAAAAAAAAAAAAAA 510				

RESULT 15	
BC048876	
LOCUS	BC048876 1901 bp mRNA linear VRT 19-MAR-2003
DEFINITION	Danio rerio, clone MGC:55559 IMAGE:2642817, mRNA, complete cds.
ACCESSION	BC048876
VERSION	BC048876.1 GI:29124507
KEYWORDS	MGC.
SOURCE	Danio rerio (zebrafish)
ORGANISM	Danio rerio
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
	Cypriniformes; Cyprinidae; Danio.
REFERENCE	1 (bases 1 to 1901)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (17-MAR-2003) National Institutes of Health, Mammalian